VOL. XVII AUGUST, 1933 No. 4

The Journal of APPLIED PSYCHOLOGY

JAMES P. PORTER, EDITOR
Ohio University
WITH A BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

E. SALNER and H. H. REMMERS. Affective Selectivity and Liberalizing Influence of College Courses 349

ARNOLD H. HILDEN. A Comparative Study of the Intelligence Quotient and Heinis' Personal Constant 355

MARION ELDERTON. An Experiment in Map Scoring and Mental Imagery Tests 376

ADELBERT FORD. Motion Analysis by the Use of Stereoscopic Motion Pictures 407

S. L. and L. C. PRESSEY. A Study of the Emotional Attitudes of Indians Possessing Different Degrees of Indian Blood 410

WALTER CROSBY EELLS. Mental Ability of the Native Races of Alaska 417

RALPH A. FRITZ. Predicting College Marks and Teaching Success for Students in a Teachers College 439

RALEIGH M. DRAKE. The Validity and Reliability of Tests of Musical Talent 447

M. E. BROOM and JOSEPHINE D. BRAMKAMP. A Study of Factors Contributing to School Citizenship 459

KATHARINE T. OMWAKE and MARGARET LORANZ. Study of Ability to Wake at a Specified Time 468

CLAUDE L. NEMZEK. The Comparative Constancy of Stanford-Binet and Herring-Binet IQ's 475

HELEN S. SHACTER. Intelligence as a Causal Factor Determining Differences in Sustained Attention in Preschool Children 478

Book Reviews 489

New Books and Pamphlets 492

PUBLISHED BI-MONTHLY BY THE EDITOR
Ohio University, Athens, Ohio
Printed by
C. E. PAULEY & COMPANY, INC.
225 N. New Jersey St., Indianapolis, Indiana

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry as second-class matter at the post office, Indianapolis, Indiana.

COPYRIGHT 1933, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. BOOK, Indiana University, W. V. BINGHAM, Personnel Research Federation, New York; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston, H. L. HOLLINGWORTH, Columbia University; J. E. W. WALLIN, Atlantic University; HERBERT S. LANGFELD, Princeton University; JAMES H. LEUBA, Bryn Mawr College; HENRY M. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, Boston, Mass.; EDWARD K. STRONG, JR., Leland Stanford Jr. University; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; JOSEPH JASTROW, New School of Social Research, New York; G. M. WHIPPLE, Danvers, Mass.; HELEN THOMPSON WOOLEY, Teachers College, Columbia University.

The scope of THE JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this JOURNAL; in every case the year, volume and pages should be specified.

THE JOURNAL OF APPLIED PSYCHOLOGY is published bi-monthly. The subscription price is $5.00 a year. Single numbers $1.00 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

AFFECTIVE SELECTIVITY AND LIBERALIZING INFLUENCE OF COLLEGE COURSES

E. SALNER and H. H. REMMERS
Purdue University

In 1931 A. J. Harris, H. H. Remmers, and C. E. Ellison made a study of the relation between liberal and conservative attitudes in college students, and other factors.[1] Their plan in brief was submitting Harper's "Social Study"[2] and a questionnaire to the members of the Elementary Psychology class at Purdue University. The Harper test, which constituted the criterion of liberality, consists of seventy-one statements to be marked in agreement or disagreement.

The students were instructed not to sign their names to either the questionnaire or the test but when these forms were completed they were asked to sign them for purposes of investigation.

The study by Harris *et al* established a mean of 47.16 with a standard deviation of 7.74 for the group of 307 students. This corresponds substantially to Harper's mean scale-score of 47 for educators holding a bachelor's degree only, in the middle and far west.

They further found from an analysis of the questionnaire a negative correlation between liberality and regularity of church attendance and no significant differences

[1] Harris, A. J., Remmers, H. H., and Ellison, C. E. "The Relation Between Liberal and Conservative Attitudes in College Students and Other Factors," *Journal of Social Psychology*, Vol. III, No. 3, August, 1932, pp. 320-336.

[2] Harper, Manly H. "Social Beliefs and Attitudes of American Educators," *Columbia University Contributions to Education, Teachers College Series*, 294, 1927.

in liberality among religious denominations or between major political parties. Also, those who did not go to church or who believed in evolution or were politically independent were more liberal than the conforming groups. There was a slight, not completely reliable, difference in liberality between men and women in favor of the former.

To conclude their study the Harper test was given to a group of 23 juniors and seniors nearing the end of a course in sociology. They found, as Harper did, a significant increase in liberality and attributed it to training or selection or both.

This question of the effect of special courses on student liberality motivated the present study. One hundred twelve students in Sociology were given the Harper test during the first week of the semester. Elementary Sociology at Purdue is an elective which draws junior and senior students from all schools (Engineering, Science, Agriculture, Home Economics, Physical Education) of the University. The test was administered in the same manner as in the previous year.

The Sociology class under consideration was decidedly more liberal at the beginning of the semester than were the sophomores of the previous study, the difference between the means of the two groups being completely reliable,[3] as shown in the following table.

	No.	Arith. Mean	S.D.	$S.D._M$	Diff $S.D._{diff}$
Sociology (Jrs. and Srs.) 1931-32	112	52.02	8.23	.78	5.41
Psychology (Sophs.) 1930-31___	307	47.16	7.74	.44	

With regard to the questionnaire items it was found that the Sociology classes did not differ to any marked

[3] Garrett, H. E. "Statistics in Psychology and Education," pp. 128-136.

extent in tendencies from the Psychology group except that when men and women were compared the women were found to be the more liberal, although the difference was not statistically reliable. This was also in contrast to Harper's results for men and women. It is possible that the course in Sociology selects more positive liberal women than it does liberal men.

At the close of the semester the Harper test was resubmitted to the Sociology group along with the final examination. The name on the examination paper made it possible to identify each test. A definite rise in liberality score was found between the beginning and the end of the semester.

	No.	Arith. Mean	S.D.	$S.D._M$	Diff[4] $S.D._{diff}$
Beginning	111	52.02	8.23	.78	
					4.58
End	112	57.09	8.41	.79	

The average for the Sociology group at the beginning of the semester (52.02) agreed rather well with Harper's norm of 50 for graduate educators in general. The score at the end of the semester (57.09) was lower than Harper's score of 68 for graduate educators after the completion of *certain*[5] courses of study.

Fifty-one students voluntarily signed their names to the test at the beginning of the semester. The Pearson correlation between their scores at the beginning and their scores at the end of the semester was found to be .69±.05. The true correlation is much higher than this, however. When corrected for attenuation[6] the coefficient becomes .90. The average score of this group

[4] *Ibid.* By the formula for correlated measures, p. 286.

[5] Italics Harper's.

[6] *Ibid,* pp. 211-213. The reliability of the Harper test is approximately .77 as determined by Harris, Remmers, and Ellison.

of students at the beginning of the semester was 54.20 with a standard deviation of 8.17, and at the end of the semester their average was 59.00 with a standard deviation of 8.84. This difference is completely reliable,[7] as is evident from the following table.

		No.	Arith. Mean	S.D.	S.D.$_M$	Diff S.D.$_{diff}$
51 before	r = .69	51	54.20	8.17	1.14	
51 after		51	59.00	8.84	1.24	5.06

The previous study of 1931 by Harris, Remmers, and Ellison established a correlation of .29±.05 (men) and .09±.05 (women) between liberalism scores and percentile scores on the American Council on Education Psychological Examination. This corresponded very closely to Symonds'[8] result of .28 between intelligence and liberalism, and to Allport's[9] correlation of .21 between radicalism and college grades.

As a further inquiry a correlation was obtained in the present study between the students' scores on four classroom tests and the final examination in the Sociology course on the one hand and their Harper test scores, and also between intelligence (American Council on Education Psychological Examination percentiles) and Harper test scores. The four tests and the examination referred to were objective tests of approximately eighty-five questions each covering the subject matter of the text book. The tests and the examination consisted of about two-thirds information questions concerning sociological problems and about one-third what might be termed

[7] Garret, H. E., pp. 128-136.

[8] Symonds, P. M., "A Social Attitude Questionnaire," *Journal of Educational Psychology*, 16, 1925, 316-322.

[9] Allport, G. W., "The Composition of Political Attitudes," *American Journal of Sociology*, 35, 1929, 220-238.

"attitude" questions. The intelligence ranking was from freshmen placement tests taken several years previously.

The correlation between grades on the objective tests and Harper scores was found to be .73±.03 for one hundred students, and between Harper scores and intelligence for eighty-nine students it was found to be .45±.05. The correlation between the objective test scores and intelligence for eighty-nine students was .55±.05. When corrected for attenuation[10] these correlations rise to .85, .54, and .60 respectively. These correlations are at least sufficiently high to indicate substantial positive relationships. It appears that the Harper test may possibly be included in courses in Sociology as a further measure of student attainment, if liberalizing the student be admitted as one of the course objectives, and that intelligence may be a factor of importance in determining an individual's liberalism or conservatism. In this connection it may be of interest to note that in answer to the question, "What magazines do you read?" those students who mentioned *Harper's Monthly, The Nation, The American Mercury,* and such periodicals had a slightly higher average liberality score than those who mentioned *Liberty, Colliers, Saturday Evening Post,* various trade journals, or none at all, there being about 70 chances in 100 of a true difference between the averages.[11]

In general the study seems to point to a possibility of measuring the so-called "intangibles" in college courses such as changes in ideals, attitudes, or beliefs, which students are generally presumed to assimilate. If a course in Sociology definitely effects an increase in liberality that is measurable it is possible that similar changes due to a study of History, Economics, English, Psychology, and the like could be treated in the same manner if suit-

[10] The reliability of the class tests was not actually determined, but was (rather generously, we believe) estimated at .90.

[11] Garret, H. E. Table XIV, p. 134.

able measuring devices were devised and utilized. It is planned to carry out such studies in one or more of these subjects in the near future.

The following general conclusions were drawn from the study:

1. The Sociology course at Purdue University operates as a selective factor in that it attracts the more liberal student.
2. The course effects a reliable increase in the liberality of those taking it.
3. While there is a significant average gain in liberality, the close correlation of initial and final measures shows that students retain their relative positions throughout. The more liberal at the beginning of the course are still the more liberal at the end and the less liberal at the beginning are also still the less liberal at the end.
4. Objective test scores in Sociology courses are possibly indicative of the liberality of the students taking them.
5. Intelligence as measured by tests may be a factor of importance in determining liberality.
6. Means may possibly be devised to measure definitely the "intangibles" in other college courses in which liberalization might reasonably be presumed to be one of the aims.

A COMPARATIVE STUDY OF THE INTELLIGENCE QUOTIENT AND HEINIS' PERSONAL CONSTANT

ARNOLD H. HILDEN

Division of Research, Minnesota State Board of Control

Is the IQ constant? This is a question that has attracted attention of psychologists since it was adopted as a method of grading intelligence by F. Kuhlmann[1] and William Stern[2]. Only insofar as it is constant does it have predictive and classificatory value. If it is not constant, does it vary in a consistent manner for different classifications of persons? Can a more satisfactory expression of the rate of mental development be found?

The problem to be considered here is the accuracy of the IQ as shown on re-tests of higher-grade subnormals from six to thirteen years of age. These results will be compared with the PC, or Personal Constant, based on the law of mental development formulated by Heinis[3].

We are indebted to Binet and Simon for the development of the concept of mental age and its adoption as a unit of mental measurement. The type of classification made by Binet was based on a comparison of the difference between the mental age and the chronological age. It was not long before investigators observed that the sig-

[1] Kuhlmann, F., "Degree of Mental Deficiency in Children as Expressed by the Relation of Age to Mental Age," *J. Psycho-Asthenics*, 1913, 17, 132-144.

[2] Stern, William, *The Psychological Methods of Testing Intelligence.* Baltimore: Warwick and York, 1914, translated by G. M. Whipple.

[3] Heinis, H., "A Personal Constant," *J. Educ. Psychol.*, 1926, 17, 163-186.

nificance of a year of retardation varies with age. This conclusion was based on actual data as well as the common-sense observation that a difference of a year is more easily seen in younger than in older children.

The next step was the adoption of the IQ. A great deal has been written about the constancy of the IQ. Since many of the investigations reported are based on group tests or deal with problems other than the constancy of the IQ, discussion is confined to those results in the literature available that deal with the problem under consideration.

At first glance the prevailing opinion seems to be that the IQ is constant. However, a closer study reveals several defects. For instance, perfect correlation does not mean that the IQ is constant. It *may* be. Also, it may not be. If the intervals between tests were the same, the increase or decrease in IQ the same, the coefficient of correlation would still be unity. Neither is it sufficient to state that the average variation, *disregarding signs*, has a certain magnitude. These two words are significant. For example, if the middle fifty per cent of differences between the IQ's on the first and second test fell between −6 and +2, the average deviation, disregarding signs, would be about 4, whereas the median difference would be about −2. The former shows the variability but does not disclose the trend of change. That is shown by the central tendency. Errors of measurement are of two kinds: chance errors and variations due to some definite trend. Often where a small positive or negative change is given no mention is made of what significance this may have for a long period of time. A change of a point or two for the group as a whole over a small period of time may not seem significant; yet, if this is a consistent change over a period of time, the final result will be marked. Sometimes several different scales have been used in the same investigation. Quite often the number of cases is small. Sometimes where as many as four or five tests have been given to many

of the individuals each result is compared with each of the others, giving a large number of comparisons but also introducing the possibility that these cases may be unduly weighted in the results. Some of the groups have individuals whose ages are beyond twenty years. This brings a complicating factor into the results insofar as the age taken as the point for the cessation of mental development may be in error.

While a number of the investigations reported are open to question or are so reported that it is difficult if not impossible to analyze them to determine any central tendency in variation that may exist, the IQ *may* be called constant, as it so often is, *if* constancy be defined as a variation of not more than five points on the average. Considering the fact that the average interval has been relatively short, this cannot be interpreted as a denial of the possibility that definite trends in variation may actually exist and become apparent over a long period of time for some levels of intelligence. Time will be taken for only a reference to a few of the reports. That the IQ is constant is the opinion of Poull[4], Wallin[5], Gordon[6], Irwin and Marks[7], Hildreth[8], Ruggs[9], Terman[10], Burt[11], and Gray and Marsden[12], altho some

[4] Poull, L. E., "Constancy of the IQ in Mental Defectives," *J. Educ. Psychol.*, 1921, 12, 323-324.

[5] Wallin, J. E. W., "Results of Retests by Means of the Binet Scale," *J. Educ. Psychol.*, 1921, 392-400.

[6] Gordon, Kate, "Some Re-tests with the Stanford Binet," *J. Educ. Psychol.*, 1922, 13, 363-365.

[7] Irwin, E., and Marks, L. A., *Fitting the School to the Child*, 1924.

[8] Hildreth, Gertrude, "Stanford Binet Re-tests of 441 Children," *Ped. Sem.*, 1926, 33, 365-386.

[9] Ruggs, L. S., "Retests and the Constancy of the IQ," *J. Educ. Psychol.*, 1925, 16, 341-343.

[10] Terman, L. M., *The Intelligence of School Children*. Boston, Mass.: Houghton Mifflin Co., 1919.

[11] Burt, Cyril, *Mental and Scholastic Tests*. London County Council, 1921.

[12] Gray, P. L., and Marsden, R. E., "The Constancy of the IQ,"

of them are open to serious criticism, especially the first three and the last one. On the whole there appears to be more evidence in favor of some tendency toward variation. That the IQ decreases with age for subnormals is the conclusion of Berry[13], Minogue[14], Prouty[15], Chipman[16], Garrison[17], Meta Anderson[18], Kuhlmann[19], and Henmon and Burns[20]. That, among the subnormals, the higher grades lose more than the lower is the opinion of Henmon and Burns, Chipman, Meta Anderson, and Kuhlmann. That the IQ's of superior children tend to increase with age is the finding of Garrison, Baldwin and Stecher[21], and can be concluded from the data given by Carroll and Hollingworth[22].

Brit. J. Psychol., 1923, 13, 315-324; 1924, 15, 169-173; 1926, 17, 20-26.

[13] Berry, C. S., "The IQ's of Mentally Retarded Children," *School and Society*, 1923, 17, 723-129.

[14] Minogue, Blanche, "The Constancy of the IQ of Mental Defectives," *Mental Hygiene*, 1926, 10, 751-758.

[15] Prouty, Ruth A., "Psychological versus Clinical Diagnosis," *Psychol. Clin.*, 1930, 18, 213-220.

[16] Chipman, Catherine E., "The Constancy of the IQ of Mental Defectives," *Educ. Res. Bull.*, 1929, 13, 103-111.

[17] Garrison, S. C., "Additional Re-tests by Means of the Stanford Revision of the Binet-Simon Scale," *J. Educ. Psychol.*, 1922, 13, 307-312.

[18] Anderson, Meta, "A Study of the Data of the Results Gathered from Repeated Mental Examinations of 200 Defective Children Attending Special Schools Over a Period of 8 Years," *J. Appl. Psychol.*, 1923, 7, 54-64.

[19] Kuhlmann, F., "Results of Repeated Mental Re-Examinations of 639 Feeble-Minded Over a Period of Ten Years," *J. Appl. Psychol.*, 1921, 5, 195-224.

[20] Henmon, V. C., and Burns, H. M., "The Constancy of the IQ with Borderline and Problem Cases," *J. Educ. Psychol.*, 1923, 14, 247-250.

[21] Baldwin, B. T., and Stecher, L. T., "Mental Growth Curves of Superior and Normal Children," *Univ. Iowa Stud., Stud. in Child Welfare*, Vol. 2, No. 1, 1922.

[22] Carroll, H. A., and Hollingworth, L. S., "Systematic Error of

The results of Meta Anderson and Kuhlmann are worthy of further attention.

Meta Anderson in a study of 200 defective children attending special classes concludes in part: "(3) In cases whose IQ is below 50, there is a tendency for the median IQ to remain constant or to gain or lose slightly. The median IQ shows a gain among the younger cases. (4) All cases with IQ's between 50 and 80 show a decrease in median IQ after repeated examinations. (5) In the case of examinations of varying intervals, the greatest loss in median IQ was at the interval of five years, which was the last examination for all grades of cases. The greatest loss in median IQ was shown by the highest grade (i. e., highest grade subnormals) of cases. (7) No group showed a marked increase in median IQ."

The examinations were made by the writer and two assistants, all well trained. No interval between tests was less than one year.

Probably one of the most comprehensive investigations of the IQ among subnormals was that of F. Kuhlmann[23] at the Minnesota School for the Feeble-Minded. The results of repeated mental re-tests of 639 feeble-minded over a period of ten years show that there is a tendency for the IQ's of the feeble-minded to decrease. His conclusions are in part: "The mental ages of the feeble-minded increase with age at a rate proportionate to the degree of mental deficiency." "The intelligence quotient decreases with age, and more for the higher than for the lower grades." "The rate of decline of the intelligence quotient increases some with age." (See Table 7 for a statistical summary).

the Herring-Binet Scale in Rating Gifted Children," *J. Educ. Psychol.*, 1930, 21, 1-11.

[23] Kuhlmann, F., "Results of Repeated Mental Re-Examinations of 639 Feeble-Minded Over a Period of Ten Years," *J. Appl. Psychol.*, 1921, 5, 195-224.

He then discusses possible explanations for this decrease in IQ. In his opinion, the fundamental reason is that when measured in terms of absolute units the yearly increments of mental ability decrease from year to year; that the growth curve is negatively accelerated. Values computed from a curve of this type show that not only do the IQ's of subnormals decrease but also at different rates on different levels in accordance with his results. Kuhlmann's discussion at this point will not be given because it will be dealt with in discussing Heinis' law of mental development. However, it should be noted that Kuhlmann's conclusions are identical with the propositions deduced by Heinis. These are also the same as the theoretical conclusions put forth by Dvorak[24], who constructed a hypothetical growth curve of negative acceleration and computed IQ values for superior, average and inferior levels.

In 1926, Heinis[25] published a law of mental development and a personal constant to express an individual's rate of mental development. Applying this constant to Kuhlmann's data[26], he found that where the IQ's of subnormals decreased from year to year, his proposed personal constant (which will hereafter be designated the PC) remains constant for all grades. Since then but one article has appeared in which his law is discussed, that of Riley[27]. Heinis expresses the law of mental development as follows:

$y=b(1-e^{-\frac{x}{d}})$, in which the variable "x" is the age, "b" is

[24] Dvorak, August, "The Relation of the IQ to Prognosis of Special Class Pupils," *School and Society*, 1924, 19, 736.

[25] Heinis, H., "A Personal Constant," *J. Educ. Psychol.*, March, 1926, 17, 163-186.

[26] Kuhlmann, F., "Results of Repeated Mental Re-Examinations of 639 Feeble-Minded Over a Period of Ten Years," *J. Appl. Psychol.*, 1921, 5, 195-224.

[27] Riley, Gordon L., "A Comparison of the PC and the IQ," *Psychol. Clin.*, 1930, 18, 213-220.

429, "d" is 6.675 and "e" is the base of the natural logarithm. To facilitate the solution of the problem it is necessary to assume that the point of origin of the curve is at birth. This is a purely theoretical consideration, since one could hardly assume that mental development is suddenly initiated at birth, and that prenatal life has had no influence on the mind. Setting the origin of the curve at zero makes possible the calculation of absolute units for mental measurement.

To find the number of units of mental growth attained by the average individual at a given age, substitute the age for "x" in the above equation and solve for "y." A table of such values has been published by Kuhlmann[28]. To find the PC of an individual, divide the number of units corresponding to the obtained mental age by the number of units for his chronological age.

As a basis for this study the values for "y" were computed for all ages in years and months from one month to forty years by means of logarithms to the sixth place. These logarithmic values were corrected to give a smooth series of values. It is believed that the results obtained are sufficiently accurate for any purpose for which they might be used. Using these logarithms of "y," a double-entry table of PC values to the fifth decimal place for mental and chronological ages was constructed.[29] Final values were taken only to the second decimal place, to make it comparable to such tables as "Inglis' Intelligence Quotient Values."[30] For example, let us suppose a child who is 10 years 7 months old has a mental age of 8 years 3 months. The corresponding number of units is 341.13 and 304.35 respectively. The PC taken to the second place is 89.

[28] Kuhlmann, F., "A Median Mental Age Method of Weighting and Scaling Mental Tests," *J. Appl. Psychol.*, 1927, 18, 621-626.

[29] Hilden, A. H., *A Table of Heinis' Personal Constants.* Minneapolis, Minn.: Educational Test Bureau, 1933.

[30] Inglis, A., "Inglis Intelligence Quotient Values." New York: World Book Co., 1925.

When the PC has been determined the degree of mental development at any particular age for that person can be found by multiplying the PC by the units at that age. This result can if desired be stated in terms of mental age by finding the age at that number of units. In the above case the number of units at five years is 226.16; 89 per cent of this is 201.28; the mental age is 4 years, 3 months. At 8 years the number of units is 267; the mental age, 6 years, 6 months.

After formulating the law of mental development Heinis presents the following thesis: "All the properties of the IQ are deduced from the law of mental development." He then gives the mathematical proof of the following propositions:

"1. The mental age of the feeble-minded increases with age at a rate proportionate to the degree of mental deficiency.

"2. For degrees lower than the average the IQ decreases with age, and more for the higher than for the lower grades.

"3. The rate of decline of the intelligence quotient increases with age.

"4. The IQ of cases above the average will increase with age instead of decreasing.

"5. The IQ of cases above the average increases with age and more for the higher than for the lower grades."

While the mathematical proofs (See Heinis[31]) of the above propositions will not be presented here, a few tables and graphs will be given to show more clearly the nature of the changes than an abstract statement would do.

The mental growth curve is shown in Heinis' article. Table I gives the IQ values at alternate years on PC levels five points apart. To illustrate: an individual

[31] Heinis, H., "A Personal Constant," *J. Educ. Psychol.*, March, 1926, 17, 163-186.

with a PC of 80 has an IQ of 74.5 at the age of 4 years, of 64.8 at 10 years, and 57.7 at 15 years. Table II has been confined to subnormals and gives the amount of loss in IQ in points at alternate even years for the different levels of intelligence. The values given in Table III are presented graphically in Fig. 1 to show the trend of change. The changes are shown at four-year intervals, on the various PC levels. For example, the IQ of a person with a PC of 50 has dropped 6.7 points in the interval between the ages of 4 and 8 years, 12.3 points in the interval between 4 and 12 years,

TABLE I

IQ Values on Alternate Years on PC-Levels Five Points Apart

PC	Year						
	4	6	8	10	12	14	16
10	7.7	6.8	6.0	5.4	4.8	4.3	4.0
20	15.7	14.0	12.6	11.3	10.2	9.1	8.4
30	24.3	21.8	19.5	17.5	16.0	14.6	13.3
40	33.0	30.2	27.3	24.8	22.6	20.6	18.9
50	42.5	39.0	35.8	32.8	30.2	27.8	25.2
60	52.8	49.0	45.4	41.8	38.6	35.6	33.0
70	63.8	59.7	56.0	52.0	48.8	45.4	42.2
75	68.8	65.3	61.9	58.3	54.7	51.2	47.8
80	74.5	71.5	68.5	64.8	61.2	57.7	54.2
85	80.5	78.0	75.0	72.0	68.7	65.4	61.9
90	86.8	84.8	82.5	80.0	77.3	74.3	71.1
95	93.4	92.3	91.0	89.5	87.5	85.4	83.1
100	100.0	100.0	100.0	100.0	100.0	100.0	100.0
105	106.9	108.3	110.3	112.8	116.4	121.1	128.9
110	114.2	117.1	121.9	128.3	138.9	159.8	375.0
115	121.9	127.8	135.6	149.2	178.3		
120	129.8	138.2	151.9	179.2			
125	139.0	150.4	172.1	235.6			

TABLE II

Decreases in IQ for Subnormals on Alternate Even Years on Various PC-Levels

PC	Year		
	8	12	16
10	1.7	2.9	3.7
20	3.1	5.5	7.3
30	4.8	8.3	11.0
40	5.7	10.4	14.1
50	6.7	12.3	17.3
60	7.4	14.2	19.8
70	7.8	15.0	21.6
80	7.0	13.3	20.3
90	4.3	9.5	15.7
95	2.4	5.9	10.3

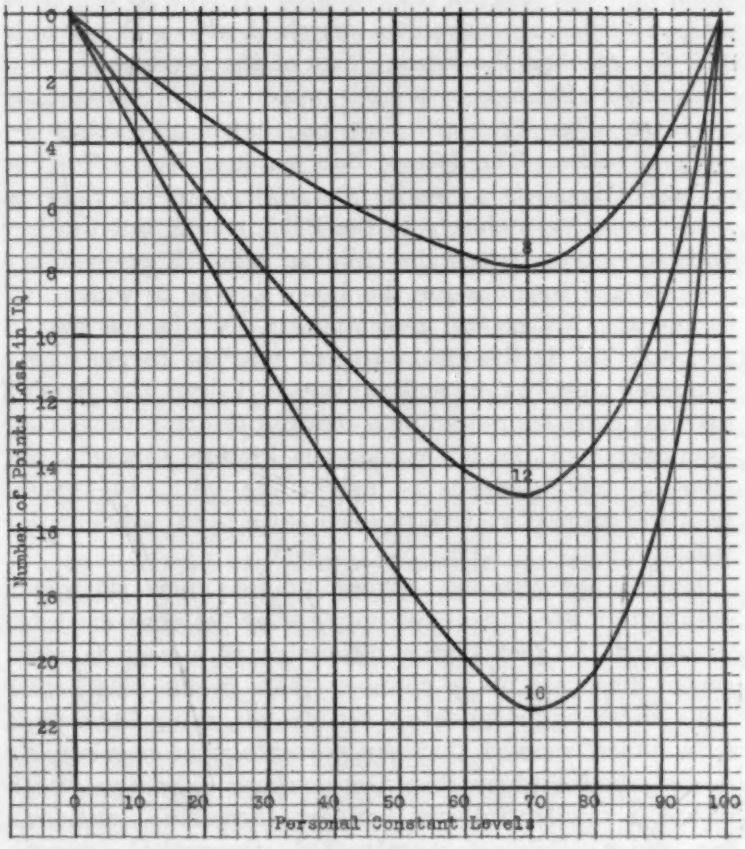
etc. The lines designated 8, 12, and 16 show the amount of change between that and the 4th year of age on the various levels; the number has been placed at the point of greatest change.

According to Proposition 2, "For degrees lower than the average, the IQ decreases with age . . . ." This is shown clearly in the graph. The graph also shows that the decrease is, ". . . . more for the higher than for the lower grades." The trend of the changes given in Proposition 3, "The rate of decline of the IQ increases with age," cannot be seen clearly. However, this is true only for the higher levels: PC 80 up to average, and up to the age of 20 years, and only to a very small degree.

Heinis concludes that the IQ's of subnormals decrease with age and more for the higher than for the lower grades. Figure 1 and Table 2 show this to be true up to the point of PC 70-75, but above that the amount of decrease becomes smaller with increase of level until the average is reached.

FIGURE I

Decreases in IQ for Subnormals



There is an acceleration of increase in IQ for all levels above the average, the acceleration increasing with increase of both age and level of intelligence. In fact, the acceleration is so marked that the question as to its significance immediately comes up. This must follow from the IQ formula and from the greater steepness of the growth curve of superior individuals: the acceleration of the numerator is so much greater than that

of the denominator. The superior child, developing at the rate of 125 per cent will advance ten years of average mental growth during the two years from 8 to 10. This accelerates as age advances. A child with a PC of only 110 has reached the mental level of the average sixteen-year-old at the age of a little under twelve years; any increase after that will not be shown adequately by present scales. The data of Baldwin and Stecher[32], (See Table VIII) show a rise in IQ for the superior children conforming very closely to the PC formula, but the last tests are given before the children were 12 years old.

Since the formula for the IQ is k=y/x, in which "k" is the rate of mental development, "y" is the mental age and "x" is the chronological age, the formula presupposed for the growth curve is y=kx. As this is geometrically expressed by a straight line the underlying assumption of the IQ is that the growth curve for all levels of intelligence is a straight line, with equal increments of mental growth for each year for any particular level.

Kuhlmann[33] among others pointed out that this assumption is contradicted not only by the general observation of the fact that a difference of a year is more easily discerned in a younger than in an older child but also by empirical data. Woodrow[34], using Bobertag's data which showed that the percentage of overlapping is greater for older than for younger children—which means a smaller interval between successive age-points for older children—constructed theoretical growth

[32] Baldwin, B. T., and Stecher, L. T., "Mental Growth Curves of Superior and Normal Children," *Univ. Iowa Stud., Study in Child Welfare*, Vol. 2, No. 1, 1922.

[33] Kuhlmann, F., "Degree of Mental Deficiency in Children as Expressed by the Relation of Age to Mental Age," *J. Psycho-Asthenics*, 1913, 17, 132-144.

[34] Woodrow, H., *Brightness and Dullness in Children*. Philadelphia: J. B. Lippincott Co., 1919.

curves of the same general type as that presented by Heinis[35].

Freeman[36] in discussing the IQ points out that if the growth curve is assumed to be a straight line, the spread of mental ages must increase from year to year. He also constructs a logarithmic growth curve in such a way as to give a constant IQ. But Kuhlmann[37] shows that Freeman apparently overlooked the fact that the growth curve he constructed does not express a constant ratio in terms of absolute units, and shows that when it *is* expressed so as to give a constant ratio of absolute units the IQ cannot remain constant. To illustrate this would be duplicating the calculating of values for Table I.

It is the fallacy of the underlying mathematical assumption that is responsible for the behavior of the IQ when interpreted in terms of absolute units of a growth curve having constantly decreasing increments with increase of age. And while the term mental age has unquestionable value for the lower levels because of its concrete and literal meaning, especially for the lay mind, it seems very probable that its use to express the development of the superior person is an unhappy one. Since a very superior person is so far above the average an attempt to express his mentality in terms of average mental levels results in confusion or in expressions which cannot be taken literally.

On the other hand inferences as to the nature of the mental growth curve should be made with caution. Nor in the final analysis is it necessary to make any. To quote from a communication of Dr. F. Kuhlmann: "The

[35] Heinis, H., "A Personal Constant," *J. Educ. Psychol.*, March, 1926, 17, 163-186.

[36] Freeman, Frank N., "The Interpretation and Application of the Intelligence Quotient," *J. Educ. Psychol.*, 1921, 12, 3-13.

[37] Kuhlmann, F., "Results of Repeated Mental Re-Examinations of 639 Feeble-Minded Over a Period of Ten Years," *J. Appl. Psychol.*, 1921, 5, 195-224.

acceptance of his (Heinis') mental growth curve is not essential to the method. It may be regarded simply as a method of weighting raw scores in such a way as to keep the derived quotients constant, and only for the particular tests we now have in the Binet scales. As these scales become more perfected the 'growth curve' may need corresponding modifications."

THE DATA

The present study is based on the data of re-tests of 433 children by members of the staff of the Division of Research, Minnesota State Board of Control. All of the examiners are trained psychologists. Hence, where both tests were not given by the same examiner it is assumed that there is a negligible difference in the results because of this factor. All examinations were made with the 1922 Revision and Extension of the Binet-Simon Scale by Dr. F. Kuhlmann.[38] These were made for various social agencies, reform schools, the Minnesota State School for the Feeble-Minded and the State Department of Education. The cases for the last named consist almost entirely of candidates for the special classes and include about 70 per cent of the cases in this study. Of 35,000 cases, about 1,900 were found with re-tests. From these were excluded all cases over 15 or under 5 years of age at the time of either test; all epileptics, cretins, etc.; all cases with mental pathology, with a language handicap or a sensory defect sufficient to make the results questionable in the judgment of the examiner; and all cases in which the emotional disturbance was marked or the response poor or to which only a part of the scale had been given. In short, in the 433 cases included in this study there are so far as is known no disturbing factors which would endanger the reliability of the tests.

[38] Kuhlmann, F., *A Handbook of Mental Tests.* Baltimore, Md.: Warwick and York, 1922.

Mention should be made of the possibility of selective factor operating to determine re-tests in some of the cases. It may be, although it is not known and cannot be determined, that some of the cases were brought in because the agency in charge felt that there was a noticeable change in the mental condition from that shown on the first test. Neither is it known whether this factor may have operated more in lowering than in raising the scores or, indeed, whether there is any relation between suspected and actual change in intellignce. This is probably the strongest criticism that can be made against the data given here.

To summarize: every attempt was made to select the results carefully with regard to the validity of both tests, with the idea in mind that a smaller group of cases selected with regard to the best conditions is much more valuable than a large group of cases, some of which might be questionable since no statistical refinement could remedy this defect.

Table III shows the distribution of the 433 cases with regard to CA, MA, IQ and PC obtained at the time of the first test. Table IV gives the distribution of intervals of time between the first and second tests.

After the final selection has been made each IQ was checked for error, then the PC for each test was recorded and re-checked. The IQ and PC obtained on the re-test were then subtracted from the respective IQ and PC obtained on the first test and divided by the interval between tests so as to get the average amount of change per year. The arithmetic mean of the change per year for the IQ's is –1.77; for the PC's –0.24. Their respective standard deviations are 3.18 and 1.99. The standard deviations in this case cannot be compared without some correction being made for the differences in the relative coarseness of the units involved. For example, let us take the number of points found between an IQ of 75 and an IQ of 100 on the twelve-year level and compare

that with the number for the PC. A child with a CA of 12 and MA of 12 has an IQ of 100, while a child with a CA of 12 and MA of 9 has an IQ of 75, a difference of 25 points. The respective values of the PC in this case are 100 and 89, a difference of only 11 points. This means that the PC has a larger unit of measurement than the IQ and that any comparison as to variability would be unfair to the IQ. The point involved here is exactly the same as that which would occur if variability in length as found in inches were compared with that found on the metric scale. Since the relative coarseness of the units varies from year to year as well as from one mental level to another, no satisfactory method was devised for equating values.

TABLE III

Distribution of 433 Cases According to CA, MA, IQ and PC Obtained on the First Test

Year	CA No.	MA No.	PC Interval	PC No.	IQ Interval	IQ No.
4			65- 69	2	45- 59	1
5	4		70- 74	9	50- 54	2
6	53	113	75- 79	21	55- 59	8
7	79	74	80- 84	39	60- 64	16
8	82	83	85- 80	110	65- 69	22
9	66	52	90- 94	201	70- 74	43
10	69	19	95- 99	50	75- 79	59
11	44	3	100-104	1	80- 84	144
12	25				85- 89	87
13	11				90- 94	41
					95- 99	9
					100-104	1
Total	433	433		433		433

TABLE IV

Distribution of Intervals of Time Between the First and Second Tests on 433 Cases

Intervals, Years and Months	No. of Cases
1-0 to 1- 5	115
1-6 to 1-11	72
2-0 to 2- 5	82
2-6 to 2-11	37
3-0 to 3- 5	42
3-6 to 3-11	27
4-0 to 4- 5	26
4-6 to 4-11	10
5-0 to 5- 5	10
5-6 to 5-11	9
6-0 to 6- 5	1
6-6 to 6-11	2
Total	433

TABLE V

A Comparison of Change in PC and IQ with Regard to Intervals Between Tests

Intervals in Years	1	2	3	4	5	6
Number of Cases	115	154	79	53	20	12
Median IQ-Change	—1.8	—4.4	—4.3	—6.2	—7.8	—7.8
Q	4.2	3.5	4.0	4.1	4.0	5.0
Median PC-Change	.0	—1.0	—0.2	—0.5	.0	+0.5
Q	2.1	2.1	2.1	2.5	3.5	3.0

A more direct method of comparison is to compare the mental age obtained on the second test with the mental ages predicted on the basis of the two mental indices. Taking the IQ obtained at the time of the first test and multiplying it by the age at the time of the second test would give the mental age to be expected if the IQ were

constant. Subtracting the mental age obtained at the time of the second test from this predicted mental age would give us a measure of variability which could be compared directly with similar values for the PC. The arithmetic mean of the difference in the mental ages based on the IQ prediction is –5.91; for the PC, –1.08. The respective standard deviations are 8.7 and 8.5 months. Since the IQ and PC are based on the same data we should expect the same amount of deviation around the mean, even though there is a difference in the central tendencies. This conclusion as to comparative variability differs from that of Riley,[39] who based his inference on differences in IQ and PC rather than on differences between obtained and predicted mental ages.

Table V shows a distinct relation between increase of IQ-change and test-interval whereas the PC is relatively constant. The small change in IQ in one year becomes considerable by the sixth year. The variability as shown by Q increases slightly with time.

Various attempts were made to analyze the data to discover any significant changes in differences on re-tests on various levels of age and intelligence to compare with values deduced from Heinis' law. There is no noticeable difference in amount of change at the various ages. However, an examination of the differences of PC levels shows that the greatest change occurs at level 94-101, the next greatest at 90-93, and the least on levels 87-89 and 68-86. This is the inverse of what is to be expected on the basis of theoretical deduction. The explanation for this difference probably lies in the possible selection of the cases at those levels. It is probable that the children with a higher mental rating who were brought in for re-examination were problem children whose lack of adjustment to society reflected nervous instability rather

[39] Riley, Gordon L., "A Comparison of the PC and the IQ," *Psychol. Clin.*, 1930, 18, 261-265.

than mental deficiency, making deterioration more probable than in what might be called the normal-dull person. Although this seems the most reasonable explanation it cannot be offered with any degree of certainty. No experimental data can be found on this point. Irwin and Marks[40] and Brown[41] find greater variability among problem children and neurotics but do not indicate any definite trend in change.

The coefficient of correlation for the IQ's is .77±.132; for the PC's, .83±.101.

CONCLUSIONS

1. For subnormals from 6 to 13 years of age, having IQ's from 50 to 100, the IQ decreases on the average 1.77 points per year.
2. The PC is a more accurate expression of the rate of mental development, the decrease for the same year being on the average 0.24 points per year.
3. The variability of the IQ and PC is about the same.
4. While the greatest decrease in IQ appears to be for PC levels 94 to 100, the possible influence of selection at this level should be kept in mind.
5. The coefficient of correlation for PC's on the first and second test is .83±.101; for the IQ's .77±.132.

The greater the interval between tests the more the IQ decreases for subnormals; the PC is practically constant.

APPENDIX

The following tables give a comparison of the PC and IQ on data which are available in such form as to make it

[40] Irwin, E., and Marks, L. A., *Fitting the School to the Child*, 1924.

[41] Brown, Andrew, "The Change of Intelligence Quotients in Behavior Problem Children," *J. Educ. Psychol*, 1930, 21, 341-350.

possible. The values in Riley's data were given by him.[42] The IQ and PC values in the Baldwin and Stecher data were calculated from their published results.[43] The values in Kuhlmann's data are not quite the same as those given by Heinis,[44] who assumed an error of six months in the mental ages published by Kuhlmann. Kuhlmann notes that there is some doubt as to the accuracy of the scores for ages 7 and 8: " . . . (corrections of) mental ages found for errors due to the tendency of the scale to measure too high or too low could not therefore be extended to low grade imbeciles, idiots or the younger morons and borderline cases."

TABLE VI
Riley's Data

Author	Cases	1st IQ	2nd IQ	1st PC	2nd PC
Doll	203	66	50	77	75
Doll	40	64	56	79	78
Baldwin	30	118	130	108	109
Riley	50	61	54	77	77

TABLE VII
Kuhlmann's Data

Age	7 ?	8 ?	9	10	11	12	13	14	15	Class
IQ	40	40	39	39	37	36	34	33	32	Imbecile
PC	53	54	55	57	57	57	57	57	57	
IQ	62	63	63	63	62	60	58	57	57	Moron
PC	74	75	78	78	79	79	79	80	81	
IQ	87	83	80	78	76	76	72	70	69	Borderline
PC	92	90	89	88	88	89	88	88	87	

[42] Riley, Gordon L., "A Comparison of the PC and the IQ," *Psychol. Clin.*, 1930, 18, 261-265.

[43] Baldwin, B. T., and Stecher, L. T., "Mental Growth Curves of Superior and Normal Children," *Univ. Iowa Stud., Stud. in Child Welfare*, Vol. 2, No. 1, 1922.

[44] Heinis, H., "A Personal Constant," *J. Educ. Psychol.*, March, 1926, 17, 163-186.

TABLE VIII

Baldwin and Stecher's Data

1. Group of 36 Children Having Five Consecutive Tests.

Test. No.	1	2	3	4	5
Av. CA	8-4	9-5	10-8	11-4	11-9
Av. IQ	115	120	120	124	127
Av. PC	106	108	107	108	108

2. Group of 42 Children Having Four Consecutive Tests.

Test No.	1	2	3	4
Av. CA	8-4	8-9	10-8	11-2
Av. IQ	111	112	115	118
Av. PC	105	105	105	106

3. Group of 51 Children Having Three Consecutive Tests.

Test. No.	1	2	3
Av. CA	8-5	9-9	10-10
Av. IQ	111	112	115
Av. PC	105	105	105

4. Group of 56 Children Having Two Tests.

Test. No.	1	2
Av. CA	7-7	8-10
Av. IQ	112	116
Av. PC	106	107

AN EXPERIMENT IN MAP SCORING AND MENTAL IMAGERY TESTS

MARION ELDERTON
University of Pennsylvania

THE PROBLEM

This experiment was devised to determine possible relationships between differences in imaginal type, as determined by certain tests of imagery and the ability to draw a familiar outline map from memory. Map drawing and the use of outline maps play an important part in the teaching of economic geography. It has been generally assumed that a person endowed with good visual imagery would, for example, fare better in terms of an efficient map performance than one whose imaginal endowment did not include such imaginal ability. At the same time, it was also desired to know what relationship existed between the grades on maps as objectively given and those given by the method ordinarily used—*i. e.*, a subjective estimate of the teacher translated into quantitative terms. This relationship should be of interest to teachers of geography where any large proportion of the work consists of map drawing.

HISTORY

The literature on tests of imagery might be said to go back as far as Fechner, who published the introspective accounts of visual imagery of several well-known persons and gave an elaborate account of his own memory images and after-images. Sir Francis Galton in his "Inquiries into the Human Faculty and Its Development" (1883) describes the results of his questionnaires espe-

cially with what he called "the visualizing faculty." Charcot came to the question from the pathological aspect. These pioneers called attention to the striking differences between individuals in the nature and abundance of particular forms of imagery. Later work, however, was especially concerned with improvement in the investigating technique.[1] Fernald carried on an intensive experiment, using eleven trained subjects, the results of which were published in 1912.

The various types of tests which have been devised have been classified into objective, subjective and a combination of the two. Fernald defines an objective test as "one which furnishes results, either qualitative or quantitative, which are sufficient to indicate an individual's type without reference to his introspection." This does not mean that such tests may not have acquired this validity by the use of the introspective data originally but simply that it is not necessary to call for introspection at present.[2] In the case of children and persons unskilled in introspection, the desirability of such tests is obvious. The purely subjective test relies exclusively on the subject's introspection, while the combination subjective-objective tests use the result of one to check on the other.

While the tests used for this investigation combine the subjective with the objective the results will be considered entirely on the basis of their objective merits; brief discussion of the subjective aspects will be included as a matter of interest only. This will therefore eliminate at least for the present the attempt to classify the subjects according to ideational types, which is not essential in making the comparisons involved in the problem set

[1] Detailed critical discussion of techniques may be found by Angell, Colvin, Segal and Fernald. See bibliography at the end of this paper.

[2] Fernald, M. R., "The Diagnosis of Mental Imagery," *Psychological Review*, Monograph Supplement, 1913, Vol. 14, No. 1, P. 13.

forth. It is perhaps a recognition of the difficulties which beset the classification into distinct imagery types,[3] and of the tendency to search instead for individual differences in images. Fernald has aptly stated the problem when she says: "It seems significant of the trend of interest in the imagery problem as a whole that the question of 'imagery types' as such has for the most part moved into the background while the search for individual differences in images and for adequate methods of determining these has persisted. This may indicate a growing suspicion that such individual differences may be more profitably stated without the incubus of 'types'."[4]

A variety of methods has been used in studying individual differences in imagery. Under what might be termed "subjective" are mainly the introspective analyses by competent observers and the questionnaire; in the "objective" category mention might be made of the memory tests with their various modes of presentation

[3] In the historical development of concepts of imagery types, there has been the swing from the belief that an individual belongs in one of three distinct groups, "visual," "auditory," or "motor" or in a fourth, which combines the characteristics of the other three, the so-called "mixed" group, to that where it was maintained that it was practically impossible to discover persons belonging to the "pure" types, and that therefore most individuals belonged in a so-called "mixed" type group, thus minimizing differences which did exist. The exact meaning of the term "mixed" was also uncertain as to whether it applied to the various kinds of imagery used in combination for one task or that a change in imagery appeared with a change in conditions. Then there was a partial return to the earlier view of distinct types. It is agreed that there are marked differences in emphasis, even though individuals use more than one image form, and the terms "visual," "auditory," etc., may be retained if they are understood to mean the predominance of a particular form of imagery. The various difficulties in the way of this seemingly simple solution are brought out by Fernald in her study. (Pp. 18-20.)

[4] Fernald, M. R., *op. cit.*, pp. 20-21.

and reproduction, with their appeal to different sensory fields; word tests, where the conclusions are drawn from the content of the words. The tests falling into the "subjective-objective" class are similar to the memory tests except for the importance given to introspection. Association tests with introspection are sometimes included under this latter heading.

The introspective analyses by the competent observers, especially where they are long continued and where the training in introspection has been adequate, are considered among the best means of studying an individual's imagery, but these conditions severely limit their application to any large number of persons—eliminating most adults and all children. Moreover, this method is uncontrolled. The questionnaire, an outgrowth of the above, seems to have originated with Galton. One of its chief limitations has been expressed by Titchener, who says: "We must remember that the questionary can never transcend or go beyond the introspections of the individual correspondents."[5]

The letter squares fall in the category of memory methods which emphasize the order and arrangement of materials. In one instance the retention of a well-defined visual arrangement, without modification or in the order learned, is called for, and in another, the reproduction in a changed order. This method is sometimes referred to as "Binet's letter squares."

The assumption underlying the reproduction of the letters of a series in an order *different* from that in which it was learned is that one who visualizes the letters will find this change easier than one who relies on auditory imagery.

Binet, investigating the imagery of the lightning calculators, Diamandi and Inaudi, used the "different order"

[5] Titchener, E. B., *Experimental Psychology, Instructor's Manual*, Qual., p. 389.

reproduction of letters in squares and suggested caution in interpreting the objective results and concluded that "the introspection of the subject, in spite of its subjective character, ought always to be taken into consideration."

Titchener used the letter squares at the Cornell laboratory in a somewhat modified form and included introspection.

Angell considered the letter square test "one of the best . . . for bringing out verbal imagery, both visual and auditory motor."[6]

Of spelling tests, considered to overlap to some degree the letter square method, Angell made the statement that the former "ought in no event to be omitted" and that "spelling backward is (even) more useful than spelling forward."[7]

Commenting on oral spelling reversing, Fernald states: "In general, we seem justified in saying that the results indicate that the rapidity of this form of spelling is favored by the ability to summon clear, vivid and stable visual images of the words without a very great need for accessory factors. A poor record, however, need not indicate an absence of visual verbal images, though it may be due to that. It may also indicate that the visual image will not stand alone, but needs other factors of the same character as those used by non-visual subjects, or that the visual image is not of a suitable character for this operation: *i. e.*, it may be of word-whole in which the letters do not stand out clearly, or it may constantly change form or disappear during the operation."[8] It is considered a partially objective test lending confirmation to the introspective reports of the subjects.

The written spelling reversing offers imagery cues

[6] Angell, J. R., "Methods for the Determination of Mental Imagery," *Psychological Review*, Monograph Supplement, 1910, Vol. 13, No. 53, p. 92.

[7] *Ibid*, p. 93.

[8] Fernald, M. R., *op cit.*, pp. 54-55.

similar to the oral, although the queer appearance of the written word tends to act as a real distraction[9] in many cases.

The foregoing will give a brief idea of the background of some of the more important imagery tests and how they have developed.

Method of the Investigation

In view of what has already been said with reference to the difficulty of classification of individuals into distinct "imagery types" and to the lack of really adequate technique to make such classification on the basis of strictly objective tests, no such attempt will be made here. The treatment of the test results will be as objective as possible, even though introspection was called for from the beginning in case it might have been of value in interpreting the individual differences. Observations of the experimenter were also noted wherever possible for whatever light they might throw on the method used by the subject. No attempt, however, will be made to offer a definite diagnosis of the subject's imagery type, partly because of some subjective factors which would thereby be injected, and partly because the purpose of the study can be accomplished without it.

A series of tests, to be described, was developed and administered to over 400 university students (most of whom were Freshmen) who were enrolled in a required

[9] The so-called "distraction principle" as a method has sometimes been used in imagery tests. It is brought out somewhat in the written spelling performance. The assumption is that "a sensory activity or an imagery process of any sort interferes with the simultaneous arousal of imagery of the same kind. Thus, sounds conflict with auditory imagery, sights with visual, movements with motor, etc. Any diminution of efficiency under conditions of distraction is taken as an evidence, therefore, that the given disturbance is allied to a favored form of imagery. The extent of the interference is considered a measure of the dominance of this imagery." (Fernald, p. 10.)

Freshman course in economic geography of the United States. In this course outline maps of the United States were constantly being used by the students as graphic aids in studying the various economic-geographical relationships. Outline maps of a certain size were usually distributed for this purpose, but occasionally the map was drawn free-hand by the student. For purposes of the experiment the drawing of an outline map of the United States was included as a part of the final examination in the course. These were graded on the basis of from 1 to 10 by the student's individual instructor—10 being the grade for the most satisfactory map.

Realizing the difficulty of securing a reliable criterion on this basis for an objective scoring of these maps they were discarded and another map was later secured from most of these students. Some of the students, however, had left the university in the meantime. For the second map each student was given a sheet of cross-section paper (5 blocks to the inch), proportioned for a correct map of the size which had been customarily used by them, with the mimeographed instructions in the center of the sheet as follows:—

> Please draw an *OUTLINE* map of the United States the maximum size that can be drawn on this sheet. Don't attempt to put in all the kinks of minor bays or capes. Remember that it is to be as large as you can possibly get on the paper. You may draw in the Great Lakes if it is easier for you to get the outline of that section in that way.

The student, of course, was not informed concerning the correctness of the proportions of the sheet although this could have been implied from the instructions. A stencil for checking the accuracy of the map was used

in scoring and the number of block deviations between the student's outline of the map and that of the stencil was the map score. Then, in order to secure a subjective (or an instructor's) score on the *same set of maps*, each map was traced by the experimenter on a plain sheet, 8½x11 inches, and the resulting carbon copy was turned over to an instructor to grade, on the basis of from 1 to 10,—10 being the best score and 1 the poorest. It was not originally intended that the instructor should use any correct sample as a guide or aid in his grading. However, due to his desire to do a careful job it was discovered that a correct sample had been employed. Because of this unintentional procedure the results are even more striking. The person scoring the maps on the subjective basis was informed as to the instructions under which the original maps had been drawn. In most instances, because of its having been transferred to a much larger sheet, the traced map had the four corner boundaries indicated to show the size of the original cross-section sheet.

Turning now to the testing materials the following problems were included in the original series:—

1. A mimeographed questionnaire, adapted from the Galton Breakfast Table Blank and the Titchener questionnaire, was used in an attempt to get some information regarding the subject's endowment and perhaps some clue as to the predominant type of imagery used.
2. Twenty (20) words presented orally to be reversed orally by the subject on which time and errors were recorded. Time was taken to tenth of a second. The observer used a mimeographed list of the oral words, in reverse order, which were easily checked as the subject spelled aloud. For example, omissions were circled, extra letters inserted, transpositions indicated, etc.
3. Twenty (20) words different from the above twenty to be written in reverse order upon the oral presentation of the stimulus word. Time was taken and recorded, as above, and errors were later checked.

4. A letter square (known as Letter Square "A") consisting of nine (9) letters arranged in three rows, and in such a way as to minimize the possibilities for association. The oral response of the subject, learning time and reproduction time in the so-called "normal" order were recorded.
5. The above series of letters was then used, with instructions to reproduce orally in the so-called "spiral" order (known as Letter Square "B"). Letters reproduced and reproduction time were recorded.
6. A letter square (known as Letter Square "C") consisting of a different set of nine (9) letters, arranged in three rows, was learned with learning time constant, and the letters given and reproduction time in the "normal" order were recorded.
7. Memory span for digits, presented visually and orally, for both forward and reverse spans.

In addition the subject's introspection, whenever possible, was secured on the Letter Square "B" ("spiral" reproduction) and on the oral and written word reversing. Observations by the examiner of the subject's overt behavior with respect to such items as lip movements, pronouncing aloud, using fingers, closing eyes, etc., were recorded where noted.

As the study progressed it became necessary to cut down somewhat on the amount of time which could be devoted to the testing of each subject. The following eliminations were subsequently effected:—

1. The oral word list was reduced by one-half,—the first ten words being retained.
2. The written word list was also reduced by one-half,—by ten words.
3. The memory span tests for digits.
4. The questionnaire.[10]

[10] The questionnaire was discarded in the treatment of results primarily because the information thus secured was of a highly subjective nature. It may be said, however, to have been of some value as it stimulated the student inexperienced in introspection to think about it. Since the questionnaire was usually distributed

All the tests were given as individual tests except the questionnaire.

The oral spelling word list was arranged so that a shorter word,—of four or five letters—alternated with the longer words in most cases. Thus in the ten-word oral list, there were two four-letter words and two five-letter words. This was done in order to ease up somewhat on the strain of spelling the longer words.

A similar arrangement held for the written list.

For the purposes of this experiment the scores on the short words have been disregarded. The score on the remaining longer words was then computed, taking time and errors into account. For example, remaining on the oral list there were six so-called long words, *i. e.*, ranging from nine to twelve letters. To the sum of the actual times taken was added a penalty of twenty (20) for each error,[11] and this figure then subtracted from 700. For example, let us say that scores were made of 15.5″, 10.2″, 7.1″, 9.2″, 18.0″, 18.0″ on the six oral words respectively, with one error,—which would result in a total score on that test of 602 (700–[78+20]). On the other hand, if say, ten errors had been made by the subject the penalty of 200 (10×20) would be added to the reproduction time (78 seconds) before subtracting from 700, resulting in this case in a score of 422. Hence a high score, *e. g.*, 602, would indicate rapid reproduction and few or no errors or a combination of speed and accuracy.

The list of words to be reversed and reproduced in written form contained five short words (four of four

to a class a week or more before the individuals in it reported for their tests, it may have helped to give the student some idea of his own imagery processes and perhaps his later introspective reportings were better than they otherwise might have been. It was not uncommon for a student to remark, "I had never thought of how I remembered things before, and since filling out the questionnaire, I've been checking up on myself to see what I do."

[11] A transposition of letters, omissions, extra letters, etc., were penalized by adding 20 for each such error.

letters each and one of five letters) and five longer words —two nine-letter and three ten-letter words. The scoring followed was similar to that for the words presented and reversed orally.

The words chosen were ones with which a college Freshman would ordinarily not have difficulty. There were no "trick" words and allowance was made for a certain type of misspelling by not penalizing the subject, if he reversed a syllable, as for example, "ence" even if the syllable to be absolutely correct was "ance." This seemed a fair thing to do since our interest was not primarily in the subject's ability to spell words correctly as such but on the reversing of what he believed to be the correct sequence of the letters in the word presented to him.

At the conclusion of the written list, which was given subsequent to the oral, an attempt was made to secure the introspection on the spelling operations.

In some cases it was hardly necessary to ask for an introspection because observance of the subject revealed the moving lips and slight throat movements used in spelling over letters or pronouncing syllables together with the halting reproduction of one, two, or three letters at a time. Sometimes spelling forward aloud was noted.

The letter square material for our experiment was divided into three parts. Letter Square "A" consisted of a card containing a square divided into nine spaces, each space containing one letter. For this series the subject was handed the card with the instructions to learn the letters across from left to right and then to turn the card over face down, when he had learned them, and to wait until he was asked before telling the examiner what the letters were.

The criterion of learning here was the honest attempt to learn the letters in order to reproduce them orally in the "normal" order when asked for by the examiner.

Approximately five seconds elapsed between the time the card was turned face down and the letters called for. The learning time, reproduction time and letters given were then recorded.

The next part known as Letter Square "B" involved this same series of letters. However, in this case the subject was asked to reproduce the letters in the "spiral" order, this order being indicated by a blank letter square with a red arrow drawn through the squares to indicate the order wanted. No further learning was permitted and the examiner entered the letters as given by the subject and recorded the time.

Letter Square "C" involving a different series of nine letters was given with the learning time kept constant, *i. e.*, 18 seconds. The subject was informed that in this instance a similar series of letters would be shown him to be learned in the order indicated (*i. e.*, the so-called "normal" order), but that there was a time limit on learning and that the examiner would turn the card over when that time was up. The subject was also assured that he would not be asked to reproduce the letters in any other order than the one learned. This was done to prevent him from anticipating a different order and perhaps thereby "wasting" his limited time on a problem not to be included. The time of reproduction was recorded for this series.

Introspection was usually requested on Letter Square "B" (the spiral order performance). It was often necessary to ask the subject when attempting to secure this introspection how he remembered the letters up the right row of the series as he proceeded around from the bottom.

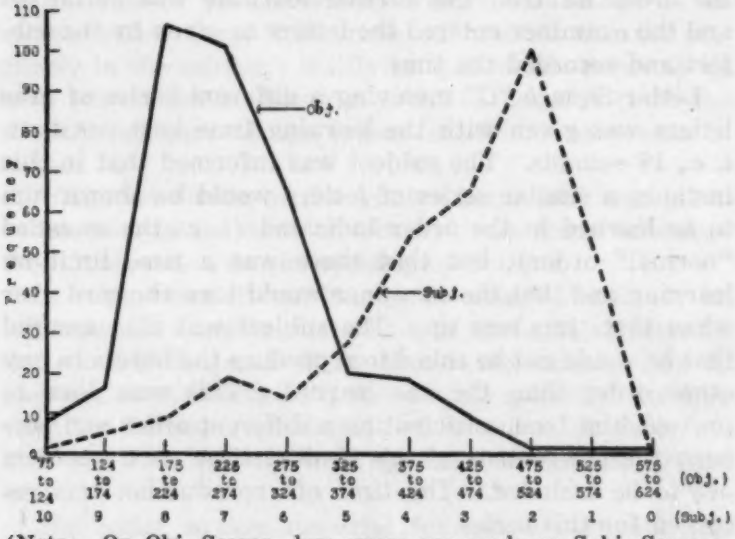
RESULTS

The results on the map scores will be given first because the data on the imagery tests are treated on the basis of grouping by map scores, which grouping will be explained and discussed later.

Figure 1 shows the distribution of the map scores both on the stencil grading and on the instructor's grading bases with the "good" scores on both laid out from the left hand side. For convenience in referring to these two sets of scores they will be designated as "objective" and "subjective" respectively.

FIGURE I

Distribution of Map Scores—Objective and Subjective Scoring Bases



(Note: On Obj. Scores—low scores are good; on Subj. Scores—high scores are good)

TABLE I

Map Scores—Objective and Subjective Bases

	Objective	Subjective
Median	242.3	3.32
Avg.	250.5±3.9	3.29±.1
S. D.	71.5	1.9
Q	44.4	1.2
V	28.5	57.5
Skewness	+.34	—.05

The combining of these results in one graph brings out clearly the significant relationship. It should be remembered that low objective map scores are good ones because they represent less deviation from the correct stencil outline, whereas on the subjectively scored maps the higher scores represent the better performance. The distribution of the objective scores is piled up at the "good" end while the line indicating the distribution of subjective map scores piles up at the "low" or poor end.

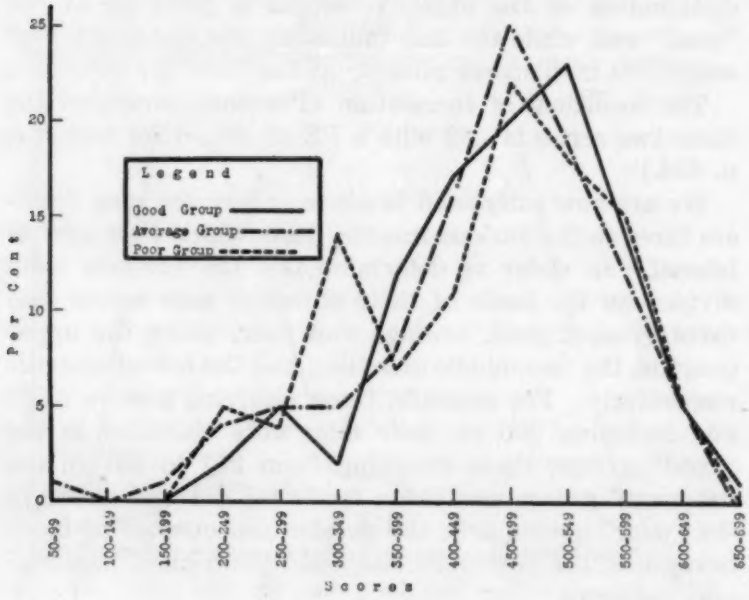
The coefficient of correlation (Pearson) computed for these two series is –.33 with a PE of .03. (See footnote, p. 404.)

We are now interested to discover how the map drawers fared in the various imagery tests which were administered. In order to determine this the subjects were divided on the basis of their objective map scores into three groups, good, average and poor, using the upper quartile, the two middle quartiles, and the lowest quartile respectively. For example, those receiving a score up to and including 206 on their map were classified in the "good" group; those receiving from 207 to 287 in the "average" group, and those receiving 288 and above in the "poor" group, *i. e.*, the greater the number of block deviations, the poorer the map and the higher, numerically, the score.

The significant measures for the various tests according to these groupings are to be found in the following tables and figures. Graphic presentation is worked out on a per cent frequency basis in order to make the results comparable. In view of the rather general negative results which characterized the data, the several charts included give a representative picture of the entire series.

FIGURE II

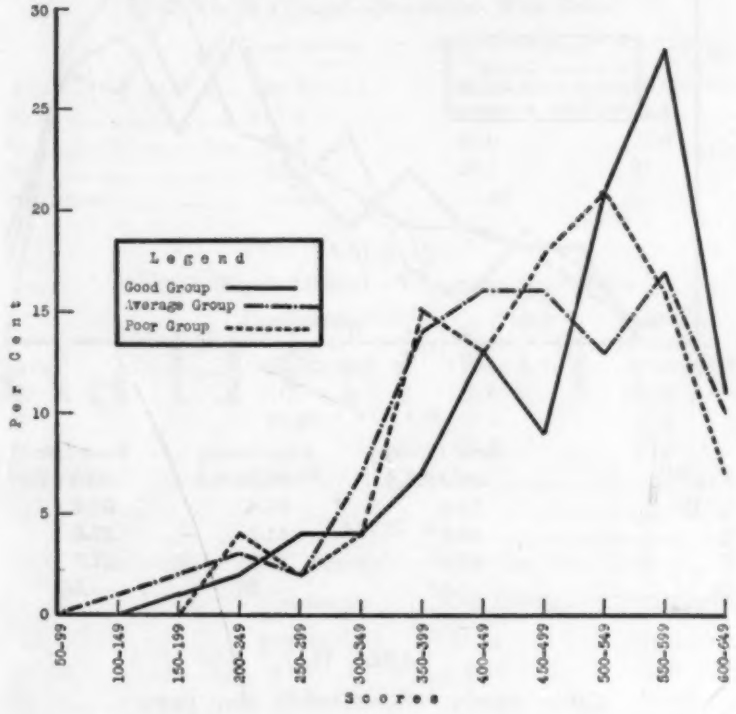
Distribution of Scores on Oral Words (Long) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	471±10.0	455.5±7.9	454.0±11.4
S. D.	93.5	103.0	105.0
Q	62.8	62.0	84.6
V	19.8	22.6	23.1
Sk	—.32	—.43	—.53

FIGURE III

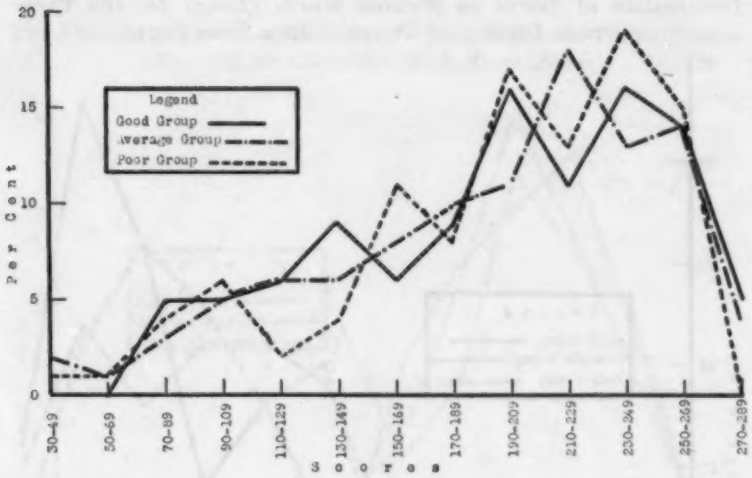
Distribution of Scores on Written Words (Long) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	494.5±11.4	459.5±8.9	470.5±1.0
S. D.	105.0	115.5	97.5
Q	73.0	85.0	74.0
V.	21.2	25.1	20.7
Sk	—.83	—.20	—.38

FIGURE IV

Distribution of Scores on Letter Square "B" (Spiral) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	196.4±5.8	193.2±4.5	195.8±5.8
S. D.	54.6	57.6	54.4
Q	43.7	41.3	37.5
V	27.8	29.8	27.8
Sk	—.43	—.67	—.56

TABLE II

Letter Square "C"—Objective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	219.3±5.8	227.3±4.1	224.3±5.9
S. D.	50.8	49.8	49.8
Q	41.4	41.1	37.2
V	23.1	21.9	22.2
Skewness	+.11	—.30	—.42

Since the scoring of maps is ordinarily not done on such an objective basis but on the subjective one mentioned earlier the subjects were re-classified into three

similar groups, good, average and poor, but this time according to their "subjective" map score. Their test results on the imagery material follows.

TABLE III

Oral Words (Long)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	460.5±11.1	466.0±7.4	446.0±10.9
S. D.	101.5	97.5	99.5
Q	60.8	67.9	73.6
V	22.0	20.9	22.3
Skewness	—.43	—.26	—.71

TABLE IV

Written Words (Long)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	484.5±11.9	477.5±7.8	448.0±13.0
S. D.	110.0	101.0	119.0
Q	87.0	74.4	85.4
V	22.7	21.2	26.6
Skewness	—.49	—.40	—.34

TABLE V

Letter Square "B" (Spiral)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	201.6±5.6	194.6±4.2	179.8±6.6
S. D.	51.6	55.4	60.2
Q	38.6	35.9	50.0
V	25.6	28.4	33.5
Skewness	—.63	—.79	—.64

TABLE VI

Letter Square "C" (Normal)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	217.5±6.0	224.0±4.2	233.0±5.2
S. D.	52.8	50.5	45.0
Q	45.9	41.4	36.5
V	24.3	22.5	19.3
Skewness	—.15	—.18	—.37

Discussion

The relationship between two methods of grading outline maps is brought out in the results of the map scores in this experiment. Figure 1 (solid line) shows the distribution of the objectively scored maps to be skewed to the right (positive), which in this case happens to be the high end numerically but the poor end qualitatively, since the larger the number of block deviations, the poorer the map. A check on the last two cases reveals a real failure to follow the instructions, in both cases a rather well-proportioned map is drawn on a much smaller scale than the standard instructions called for. Adjusting a stencil for a smaller map, these two scores would be quite good. If these two cases were discarded the skewness for the distribution is changed to ±.31 as against +.34. The average is also raised qualitatively (hence lowered quantitatively) to 248.7 as against 250.5.

If the criticism be made that an otherwise good map is penalized by the use of the stencil, because of the necessity for the stencil to be definitely placed at certain points, the answer seems to be that in the ultimate deviation there would be no benefit to the score. For example, it might be claimed that in a particular map the eastern coast line is quite accurate but started over to the left too far, and by placing the stencil in its correct place to the right the "goodness" of the east coast line is not properly evaluated. Suppose then the stencil were shifted to the left to "fit" the eastern coast, then the map drawer would lose out at the other side due to the fact that the left side of the sheet limits the subject from proportionately extending the western outline and so the points gained by shifting the stencil to the left to take advantage of a good Atlantic Coast line would be offset at the other side by the fore-shortening of the Pacific Coast line because of the limitations of the paper on which the map was drawn.

It may be asked, why bother to grade maps so care-

fully when a conscientious effort on the part of the teacher can do a similar job with much less time involved? As far as the time element is concerned there is no argument although the additional time involved with a stencil is not great, but when a *similar* job is desired there may be some room for doubt. However, on a task that is expressed in such an objective manner as the outline of a map one might logically expect a rather close relationship to result from the two methods.

A most careful job was done on the marking of the duplicate set of maps with the aid of a carefully prepared copy of the size of the original map kept before the instructor and with the four corners indicated of the boundaries of the cross-section sheet.

The distribution of these scores, known as the "subjective" basis scores, is also shown in Figure I (broken line) and arranged with the best scores, qualitatively, at the left to correspond to the arrangement in the other map series. The distribution here is skewed (–.05) in the opposite direction. The piling up is at the low score (*i. e.*, poor map) end, and the mode is at 2. Superimposing the two curves on the same figure makes the lack of similarity apparent.

When the coefficient of correlation (Pearson) was computed it was found to be –.33, with a PE of .03. (See footnote, p. 404). This relationship while slight is statistically significant, being eleven times its probable error. Such a result is surprising in view of the careful, conscientious job which the writer knows was done on this set of maps. This is another example of relative unreliability of subjective teachers' grades as compared with objective criteria.

Imagery Tests and the Objective Map Results

In order to make comparisons between the results on the imagery tests and the map scores it will be recalled that the subjects were divided into three groups—the

good, average and poor—on the basis of the map score. This comparison is made first on the objectively scored maps. Spelling words backwards and letter square tests constituted the imagery tests used in this experiment.

Tests on backward spelling have been used by a number of investigators although Angell states that they seem to have been used less than one might anticipate. He suggested that they are an admirable form of test to use in conjunction with tests on descriptions inasmuch as it brings into the foreground peculiarities of verbal imagery as compared with the imagery of objects, and that this virtue constitutes its main limitation. It would seem, however, that if one does not interpret the term "verbal" too literally this test also seems to combine the "object" or concrete category with it. The task of spelling a fairly long word backward sets up the combination of letters *as an object*,—they are seen in different kinds of letters,—large, small, printed, written, fancy, black on white, and in various spacings, etc. The short four- and five-letter words scarcely presented any problem and the response was more or less automatic, so that they were excluded in treating the results. They did, however, fulfill a real need in furnishing relaxation and encouragement, coming as they did between longer words. The latter to many presented a real problem and there was little opportunity for the performance to become an automatism.

One of the interesting things brought out by the introspection and which seems not to have been stressed by investigators of imagery was the importance of the motor element, even for those whose visual images were clear, easily stimulated, and relatively lasting. In nearly every case the subject admitted that before any helpful cue for reversing was obtained some pronouncing forward to himself or a spelling of the word forward to himself was the first response. This would call up either a visual image of the whole word or parts of the word,

of syllables or one or two letters at a time and if the image did not last sufficiently long, it was continuously revived and reinforced by auditory and/or motor images. Except in cases of especially good visualization, where the word could be set up visually and read off backwards at an even rapid rate, there was a constant spelling forward (or pronouncing) to the place where the subject had left off reversing. Cases were found where audible spelling and pronouncing had to be resorted to in order to do the test. It was common testimony that when the letters disappeared or faded, the subject would spell forward or pronounce again and again. Shaking the head to emphasize syllables and counting on the fingers often accompanied it. Sometimes in the struggle to hold the images the subject would become confused and not remember accurately "where he was." This would sometimes involve the insertion of extra letters or syllables or of transposition of letters. A number of records, a very small proportion, indicated the rapidity and accuracy of those whose introspections bore testimony to clear, lasting visual images. An even smaller number on this test whose auditory ability seemed highly efficient compared almost as favorably as did some of the former. In other instances one could see the lips actually moving in efforts to retain enough of the image of the word to give back, haltingly, a letter or two at a time. If you think it is an easy matter, especially if you happen not to be highly endowed with clear visual images, try to spell backward "Philadelphia" or some other equally long word. Where the vocal-auditory and visual cues were efficiently combined in the subjects a fairly creditable performance resulted.

If we take now the three groups divided on the basis of the map drawn (and objectively scored) and examine the results obtained on the oral spelling test we find (see table below Figure II) that there is little differentiation. It will be remembered that a high numerical score on the

oral word test is a good score. There is a slight difference in favor of the "good" group with an average score of 471, but there is practically no difference between the means of the "average" and "poor" groups, 455.5 and 454.0 respectively. The coefficient of variation (Pearson) for the entire group is 22.3 as compared with 19.8, 22.6 and 23.1 for the separate groupings. There is a negative skewness which increases from the good to poor groups.

The graphic representation of these data (Figure II) shows clearly the essential similarity of these groups.

The scores on the written words are shown graphically in Figure III. The written spelling operation seemed to be somewhat easier for many of the subjects because as they themselves expressed it they "could tell where they were" more readily, and did not need to spend so much time remembering where they had stopped on the reversing operation. On the other hand the queer appearance of the words written this way seemed to confuse and distract in some instances. This was brought out particularly by the introspection of those who had claimed good visual images, who insisted that the appearance of the writing "knocked the word completely out of" their minds, and it was only after they hit upon the method of writing with the eyes closed or of looking away while writing, that this distraction was removed. This testimony seems to lend some weight to the so-called "distraction principle" mentioned earlier. The effect was certainly more noticeable with that small proportion of the group who claimed great ease of visualization.

On the whole the written scores were better than the oral regardless of the extra time which was presumably involved in the manual operation of writing.

This seems to be somewhat at variance with Fernald's findings on the written spelling tests. In summarizing this section of her study she states: "In general, it is evident that the procedure is more difficult than the preceding (oral), chiefly because the peculiar appearance is

sufficiently disturbing to more than counteract any advantage that might be expected from the sight of the letters already written."[12] It can hardly be accounted for on the basis of different words, since the five longer words retained on the written spelling list were taken from the Fernald list. It might, however, be partly due to the fact that the average length of the written words was 9.8 as compared with 10.7 letters for the oral.

In the results for this test the best group continued to show the highest average, but the difference between it and others is not statistically significant. The low group is the least variable, although there is little difference in this respect when all are considered. All three curves are skewed at the high end, but the best group shows the greatest amount, as might be expected. Confirmation, therefore, on the basis of written scores of a real difference among the groups according to map drawing ability is lacking here.

In the letter square test spatial localization of the items which are used in the squares is considered as a help in contrast to the so-called "distraction principle." Angell states that it is one of the best in the list (of those his committee was evaluating) for bringing out verbal imagery, both visual and auditory-motor.

In the performance of Letter Square "B" it may be recalled the subject reproduced a series of nine letters in a spiral order, having learned them in a so-called normal order and without knowing in advance that a different order would be called for. The scoring here penalizes for errors and takes account of speed of reproduction as well as of learning, for it includes the sum of the reproduction and learning times plus the penalty of 20 for each error,—this whole sum subtracted from 300. A high score, therefore, is a good score. An error is defined as an omission, a letter substituted or not in the list or a letter not in its correct "box."

[12] Fernald, M. R., *op cit.*, p. 57.

In regard to learning time, this was a variable factor since the criterion was simply the honest attempt to learn the series so that it could be reproduced in the normal order.

A changed order of reproduction for a good visualizer, all other things being equal, has ordinarily been assumed to be a relatively easy task and one should therefore expect to find a differentiation among the three groups, assuming further, of course, that excellence of map drawing represented some relationship to the individual's dominant form of imagery.

Figure IV reveals practically no differentiation whatever and the results are almost identical with those for the group as a whole. If the headings were omitted from the columns one would hardly realize that three supposedly distinct groups were being compared.

The part played by association is difficult to evaluate. While the letters chosen were supposed to minimize this aspect of the test, it was surprising what ingenuity was resorted to in an effort to remember the letters. The most common type of association, however, was that made with initials of names. As far as the introspection was concerned it seemed that the saying over to oneself of the entire line to get the needed letter was most frequent. This process might be assumed to be accomplished quickly without any great amount of visual cues if the original learning had been carefully done. It was the rare subject to whom all the letters appeared simultaneously and remained long enough to enable him to merely read them off in the order wanted. In several instances subjects reported that the three letters on a line at a time "flashed into mind" and gave sufficient usable cues to reproduce without any further effort. To be sure these individuals were the exceptions. Most subjects, however, reported auditory-motor with visual in varying degrees. These tests as a whole tended toward

the appearance of visual memory images even though somewhat vague in many cases. It appears that the articulatory image oftentimes gives the cue for the reinstatement of the visual image when the latter has been lost.

In order to equalize somewhat the difference between the slow and rapid learners the learning time on Letter Square "C" was kept constant at 18 seconds. This amount of time set arbitrarily turned out to be about half of the average time taken to learn Letter Square "A"—computed from those cases where the reproduction was perfect in the so-called normal order. On Letter Square "C," the subject was definitely assured that he would not be asked for a reproduction in any but the order in which he had learned the letters.

Table II gives the significant results on the scores for this series. The score was computed by subtracting the sum of the reproduction time and 20 times each error from 300. It will be noted that the range of scores in this series is much narrower than in Letter Square "B." The performance of a task in an accustomed manner, *i. e.*, in the normal order, does not appear to make for any great variation.

The negative picture, however, is continued here so far as differences among the high, middle and low map groups are concerned. The middle group seems very slightly better than the other two.

Due to the part played by memory in learning the squares and the difficulty of controlling overlearning or underlearning, it would seem that the advantage in tests of this kind might be in favor of the spelling tests. Here, if the subject knows the word or if it is not too difficult, he has an opportunity to use the imagery best suited to him. However, it cannot be assumed that he will always use this procedure.

Imagery Tests and the Subjective Map Results

There would be no point in going further and making a similar analysis of the imagery test scores, re-sorting into three groups by the map scores graded on the subjective basis, were it not for the fact that the latter method of grading maps is practically the only one used, and had the other procedure not been tried out the customary method might have been considered the logical basis on which to have made the sortings.

The groups here were divided similarly to the previous three, *i. e.*, the highest quartile (good), the two middle quartiles (average) and the lowest quartile (poor). A consideration of scores on oral words spelled backwards shows that the mode is the only measure in which the good group can claim a seeming superiority, 525 against 475 for the other two groups. The mean of the average group is slightly better than that of the good group, although not significantly so. The poor group average, however, is definitely lower than the others, 446 as against 460.5 (good) and 466 (average). But when the variation is considered the groups present no clear-cut differentiation. (See Table III.)

From Table IV (written scores for words spelled backward) we get some differentiation. The averages for the groups show a decreasing tendency from the good to the poor group—484.5, 477.5 and 448—which is what one would be led to expect—but in the other measures the average group varies least from the central tendency so that here again, when the whole picture is taken into consideration, we cannot argue for any clear-cut evidence that the subject's ability displayed by his map work is closely related to his performance on this particular test.

An analysis of Tables V and VI present two nearly opposite pictures. In the former with respect to Letter Square "B," we have practically the only data which give

a consistent advantage to the good group, a lesser one to the average or middle group, and the poorest showing to the third group.

Letter Square "C" results on the other hand are almost the reverse, with the poor group carrying off the laurels for superiority, such as they are.

Referring back to the actual map drawing and the most adequate way to execute a task such as this from memory, it has been claimed that an individual who is a ready visualizer, who can "see clearly in his mind" the whole outline at a time or even parts of it, should be able to translate that picture easily and effectively into an accurate outline. On the other hand it is conceivable (and demonstrable) that a person totally deficient in visual imagery could make a very creditable showing on such a task, with the imagery cues which come from the kinaesthetic field predominating. Just consider for the moment that you have been asked to draw such a map as was called for in this experiment and suppose also that you do *not* have a very clear picture of it in your mind. Do you find yourself using auditory-motor verbal aids in locating the various points? For example, do you find yourself saying that the New England States come down like this, Cape Cod juts off here, New Jersey turns in thus, the Gulf of Mexico comes up and goes around like this, etc.? Then the instant you make a line on your sheet of paper you begin a comparison of the actual with the memorial or imaginal, shifting quickly back and forth, checking the image, and getting the cue for the next movement. It seems if one performs the task slowly and carefully that the auditory and vocomotor verbal aids are likely to be experienced and prove helpful. Whatever ability the map drawers in the best group had in the experiment it showed little differentiation when set up along side the other two groups' results on the imagery tests.

CONCLUSIONS

1. The most striking conclusion which appears refers to the method of grading the maps. From the results of this experiment, a slight although significant relationship is shown to exist between subjective and objective map scores, represented by the correlation of only –.33 (PE .03).* There is an indication that the instructor who did the scoring tended to keep the general shape of the map in mind, as indicated by this positive relationship, which of course does not have the accuracy made possible by the actual objective measurement.

2. There are no significant differences in correlation between map scores and tests of imagery among the three groups on either basis of sorting. The closest approach to consistent (but not statistically significant) differences is found in the results of the Letter Square "B" (subjective basis) where the good group is somewhat superior to the average, and the latter to the poor group.

3. A consistent inverse relationship is found to exist among the groups for Letter Square "C" (subjective basis), but it is not statistically significant.

4. Whatever abilities are involved in the drawing of the map they are not significantly related to those involved in the imagery tests as administered and scored in this experiment. The former may be much more complex and varied than was assumed.

5. At the university level, regardless of the advantages which might be obtained at lower levels, the actual drawing of maps might be dispensed with in view of the

* This coefficient is preceded by a minus sign because high scores represent good performance on one test and low scores good performance on the other.

low degree of correlation between the methods of scoring the maps.

6. It might be interesting if a satisfactory technique were evolved for classifying the subjects into four groups on the basis of the introspection and combined score computed for all the tests into a so-called (1) good visual, (2) intermediate (vis.–aud.–mot.), (3) poor visual and (4) good auditory. Then a comparison could be made with the maps which would represent an opposite approach to the one which has just been made. This kind of classification necessarily emphasizes the subjective aspect but the introspection would have to be used as a check for the objective test record.

7. The paper presents new objective methods for both the scoring of drawn maps and for the combination of times and errors in a single score for the tests of mental imagery.

BIBLIOGRAPHY

(Note: Only a few selected references will be included here, chosen primarily because of their historical significance. The bibliography contained in Fernald's study contains 118 references, which may be consulted should more material be desired.)

1. Angell, J. R., "Methods for the Determination of Mental Imagery," *Psychological Review*, Monograph Supplement, 1919, Vol. 13, No. 53, 61-107.

2. Bell, A., and Muckenhoupt, L., "Comparison of Methods for the Determination of Ideational Type," *American Journal of Psychology*, 1906, 17, 121-126.

3. Binet, A., *Psychologie des grands calculateurs et joueurs d'échecs.* Paris, 1894.

4. Charcot, J. M., *Leçons sur les maladies du système nerveux. Oeuvres Completes*, 1886-1890, 3, 154-193.

5. Colvin, S. S., "Methods of Determining Ideational Types," *Psychological Bulletin*, 1909, 6, 223-237.

6. Fechner, G. T., *Elemente der Psychophysik*, 2, 469 ff.

7. Fernald, M. R., "The Diagnosis of Mental Imagery," *Psychological Review*, Monograph Supplement, 1912, Vol. 14, No. 1.

8. Galton, F., *Inquiries into the Human Faculties.* London, 1883.

9. Galton, F., *Statistics of Mental Imagery. Mind,* 1880, 5, 301.

10. Segal, J., "Ueber den Reproduktionstypus und das Reproduzieren von Vorstellungen," *Arch. für das ges. Psychol.,* 1908, 12, 124-236.

11. Titchener, E. B., *Experimental Psychology, Student's Manual.,* Qual., 195-200; *Instructor's Manual,* Qual., 387-402.

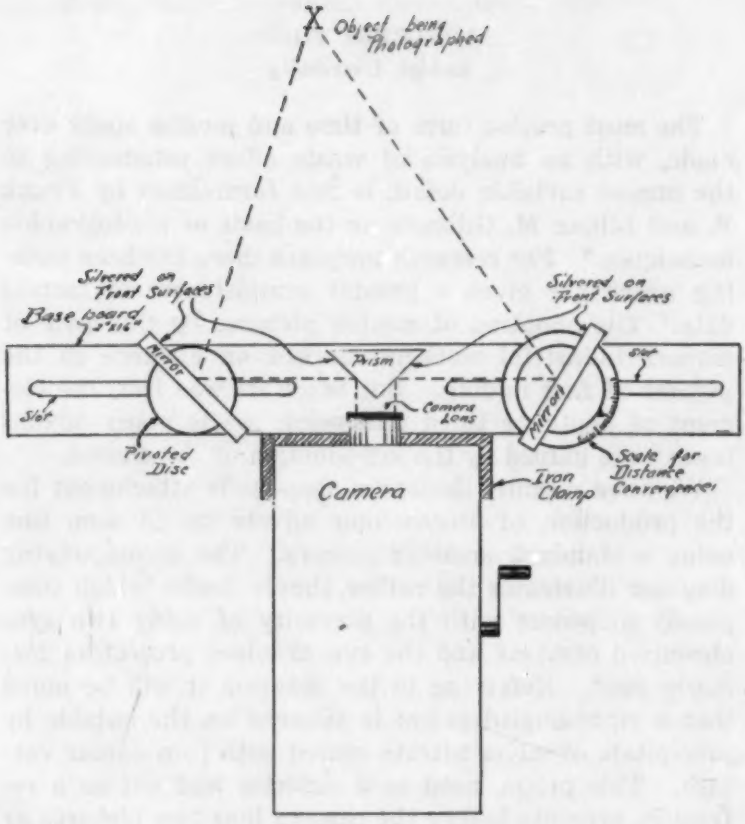
MOTION ANALYSIS BY THE USE OF STEREOSCOPIC MOTION PICTURES

ADELBERT FORD
Lehigh University

The most precise form of time and motion study ever made, with an analysis of waste effort penetrating to the utmost invisible detail, is that formulated by Frank B. and Lillian M. Gilbreth on the basis of photographic techniques.* For research purposes there has been nothing which has given a greater completeness of factual data. The adoption of motion pictures by this pair of pioneer industrial workers marked an advance in the process of fact finding. But one item was lost, the element of depth or third dimension, while many advantages were gained by the introduction of the cinema.

We have recently devised a shop-made attachment for the production of stereoscopic effects on 16 mm. film using a standard amateur camera. The accompanying diagram illustrates the rather simple device which completely dispenses with the necessity of using two synchronized cameras and the synchronized projectors formerly used. Referring to the diagram it will be noted that a right-angled prism is silvered on the outside by precipitate of silver nitrate coated with thin damar varnish. This prism, used as a reflector and not as a refractor, presents before the camera lens two pictures or images which appear side by side in the film "frame." On the mounting board which is clamped to the front of the camera are placed two plate glass mirrors, silvered on the front surface to prevent double reflection, at such an

* F. B. Gilbreth, *Applied Motion Study*. New York, 1919.


Object being Photographed
Silvered on Front Surfaces
Base board 1"x16"
Mirror
Prism
Camera Lens
Slot
Pivoted Disc
Camera
Iron Clamp
Scale for Distance Convergence

angle that one of them presents the view of the object as seen by the right eye and the other the view as seen with the left eye, each image being carried into the camera by one of the faces of the right-angled prism.

For distant objects an increase in the depth effect is often desired and can be secured by widening the distance between the plate glass mirrors, giving the same effect as secured by the widening of the inter-pupillary distance in prism binocular field glasses. The slots in the base board allow for this adjustment. One of the plate glass mirrors is made adjustable by the use of a thumb screw and rotating bearing so that the image may be centered on either near or far objects. A scale indicates the position without the use of a special view finder.

In viewing these pictures the investigator, presumably motion study expert, mounts the projector behind a ground glass screen placed in the rack of an old style Brewster stereoscope such as was found a few years back on every parlor table. The distance of the projector is so adjusted that each of the pair of pictures shows a cross diameter of about three and one-half inches. By the use of special plane prism viewers these pictures may be shown on a screen in the ordinary way.

A STUDY OF THE EMOTIONAL ATTITUDES OF INDIANS POSSESSING DIFFERENT DEGREES OF INDIAN BLOOD

S. L. and L. C. PRESSEY
Ohio State University

As a by-product of various investigations into the differences between Indians and Whites there have appeared articles[1] showing the variations in results obtained from Indians having more or less admixture of white blood. In general these articles have reported a gradual but fairly significant change in the test scores obtained from the full-bloods as compared with breeds who possess less and less Indian heritage. The typical situation seems to be that the more white blood a group possesses the less they resemble full-blooded Indians and the more they tend to approach the norms of the full-blooded Whites. Similar findings have been presented for the Negroes and various oriental groups.[2] All this is quite as the eugenieist would have it.

[1] See for instance, Garth, T. R., "The Intelligence of Mixed Blood Indians," *Journal of Applied Psychology*, 1927, XI:268-275; Jamieson, E. and Sandiford, P., "The Mental Capacity of Southern Ontario Indians," *Journal of Educational Psychology*, 1928, XIX: 313-328 and 536-551; Paschal, F. C. and Sullivan, L. R., "Racial Influences in the Mental and Physical Development of Mexican Children," *Comparative Psychology Monographs*, 1925, III:1-76.

[2] See for instance Gray, C. T. and Bingham, C. W., "A Comparison of Certain Phases of Musical Ability of Colored and White Public School Pupils," *Journal of Educational Psychology*, 1928, XX:501-506; Young, P. C., "Intelligence and Suggestibility in Whites and Negroes," *Journal of Comparative Psychology*, 1929, IX:339-359; Porteus, S. D., Dewey, D. M., and Bernrenter, R. G., "Race and Social Differences in Performance Tests," *Genetic*

During the current year the writers have been investigating certain emotional attitudes among rural white children and Indian children.[3]. They have used a scale (a much-improved revision of their earlier X-O examination[4]) that calls for pupils to single- or double-cross (according to the intensity of their feelings) those things they regard as wrong, those things they worry about, those things they are interested in, and those characteristics they admire in people. For each of these expressions of attitude 90 items are presented, giving a total of 360 in the examination—called the "Interest-Attitude Tests." These items are scored differentially in such a way as to give the best possible indication of maturity. In fact, the tests form essentially a scale of emotional maturity. The scores may be kept in terms of points (in which case they may run from −200 to +520) or they may be expressed in terms of a carefully worked-out "emotional-age."[5] The writers, in the present article, have used the emotional age rather than the raw score because of reasons which will presently appear, but its use in no way affects the nature of the results. The scores made by full-blood Indians and various types of breeds would appear in precisely the same relation to each other in any case; expressing these scores as emotional ages merely permits of a more convenient handling, without altering the significance of the data in the slightest degree and without introducing any artificial or irrelevant features.

Psychology Monographs, 1930, VIII:95-208; Louttit, C. M., "Test Performance of a Selected Group of Part Hawaiians," *Journal of Applied Psychology*, 1931, XV:43-52.

[3] Under grant from the National Research Council.

[4] Pressey, S. L. and L. C., "Development of the Interest-Attitude Tests."

[5] Anyone desiring more information than is given here concerning the technique of establishing the emotional ages used in this study should consult the previous paper (referred to in note 4) where the matter is gone into at some length.

The cases that have been tabulated for the present study[6] consist of 216 Sioux, 2 Assiniboine, 64 Omahas, 47 Pottowatomies, 84 Winnebagos, 122 Cherokees, 52 Arapahoes, 43 Cheyennes, 126 Crows, 140 Utes or Paiutes, 29 Shoshones, 6 Comanche, 5 Chemehuevi, 30 Creeks (Choctaws or Chickasaws), and a number of Californian tribes (totaling 366 cases from 22 tribes). In all these tribes there is a considerable admixture of white blood.[7] All cases tested were classified as being full-bloods, three-quarter breeds, half-breeds, or quarter-breeds; such fractions as 5/16ths, 1/8th, 15/16ths of white blood, and so on, were included in the nearest group. Information on this matter was obtained from school records. When all the data had been gathered together the following number of individuals appeared in each group.

TABLE I

Showing the Number of Children of Each Sex in Each Classification

	Full-Bloods	¾ Breeds	½ Breeds	¼ Breeds	Total
Boys	296	96	154	78	624
Girls	310	117	175	106	708
Total	606	213	329	184	1332

Because the girls have slightly different norms for reading raw score into emotional age, the results from the two sexes were kept separately. But there seemed no good reason, once the scores were in terms of emotional age, why the results should not all be combined in a single table. Results are, therefore, given in detailed

[6] Omitted were the Pueblo Indians and non-Pueblo tribes of the Southwest because the number of breeds was not over 1% and thus offered no chance for comparison.

[7] As indicated in the *Indian Population in the United States and Alaska*, by S. L. Rogers, Government Printing Office, 1915.

form only for both sexes combined, but the summary is presented for each sex separately.

The method of presentation used is merely to tabulate in adjoining columns the chronological and emotional ages of those in each classification, then to find both the median chronological and the median emotional age for each group, and finally to subtract one of these ages from the other. The results of this procedure for the entire group appear in Table II.

TABLE II

Showing the Distributions of Chronological and Emotional Ages for 624 Indian Boys and 708 Indian Girls.

Ages	Full-Bloods		¾ Breeds		½ Breeds		¼ Breeds	
	Chr.	Emo.	Chr.	Emo.	Chr.	Emo.	Chr.	Emo.
8		5		3		3		3
9		30		10		15		15
10	1	51		21	2	33	1	18
11	14	51	5	23	6	16	9	18
12	37	58	11	23	17	40	13	17
13	44	80	29	22	31	47	19	22
14	70	88	26	40	48	65	23	34
15	78	78	30	23	45	39	26	19
16	102	73	38	19	60	27	36	12
17	108	53	25	11	51	23	16	18
18	72	33	25	13	37	13	21	4
19	47	12	14	4	19	8	10	3
20+	33	2	10	1	13		10	1
Totals	606	606	213	213	329	329	184	184
Median Chronological Age	16.5		16.1		16.2		16.0	
Median Emotional Age		14.3		14.1		14.2		13.9
Years of Emotional Retardation	2.2		2.0		2.0		2.1	

The above table should be read: The median chronological age of the 606 Full-Bloods was 16.5 years; the

median emotional age of the same group was 14.3 years; the median emotional retardation was, then, 2.2 years. Similar figures for the other groups are to be read in the same manner.

For the sake of anyone interested in sex differences in emotional attitudes the following summary is presented. The figures are those obtained from detailed tables in which results were tabulated by sex; these tables were made in identically the manner described for the combined table presented above.[8]

TABLE III

Showing the Median Chronological and Emotional Ages, and Median Emotional Retardation, by Sex.

	No. Cases	Median Chronological	Median Emotional	Years of Emotional
Boys		Age	Age	Retardation
Full-Bloods	296	16.6	14.6	2.0
¾ Breeds	96	16.1	14.1	2.0
½ Breeds	154	16.0	14.0	2.0
¼ Breeds	78	16.0	13.9	2.1
Girls				
Full-Bloods	310	16.6	14.4	2.2
¾ Breeds	117	16.2	14.1	2.1
½ Breeds	175	16.3	14.2	2.1
¼ Breeds	106	16.0	14.0	2.0

The differences either for each sex or for all cases combined in (a) chronological ages, (b) emotional ages and (c) amounts of retardation among the four classes of Indians are all slight and probably not important. To be sure the breeds of all degrees are about a half year younger chronologically than the full-bloods (probably because they are somewhat brighter and tend to progress a bit faster in school) but their amount of

[8] Anyone interested in these tables may obtain them from the writers.

emotional retardation is not significantly different. Both the distributions for chronological age and that for emotional age show a dispersion of the cases over the same general range of ages for all groups. In no instance does there appear to be any real tendency for more and more admixture of white blood to result in an approach of the scores to the white norms, from which the emotional ages were derived. The Indians seem to make about the same scores regardless of the "purity" of their blood, but there is a sharp break between the whitest of the breeds and the 3,875 white children on whom the norms are based.

The data certainly suggest a triumph for the environmentalist. Apparently if a person of any degree of blood lives and behaves like an Indian he thinks and feels like an Indian—and for all social purposes, *is* an Indian. A quarter-breed may (or may not) be definitely brighter than a full-blood but he evidently tends to have emotional reactions markedly similar to those who have no white inheritance. In most if not all of the tribes included in the study, the breeds live with the full-bloods under the same conditions, go to school with them, stay with them, intermarry with them, belong to societies with them, without discrimination. One would not, then, expect that results on a scale intended to reflect acquired emotional attitudes[9] would vary significantly according to different amounts of Indian or white blood as long as children of all classifications grow up together in a similar environment and as long as nobody cares about differences in lineage. Presumably any real discrimination against the breeds or any success on their part in mingling on intimate social terms with white people should result in significant difference in scores. The writers would suggest that an interesting experiment

[9] There seems excellent evidence that these tests are not measures of intelligence. For details the reader is again referred to the article mentioned in note 4.

could be carried out on the attitudes of breeds from any race in relation to their acceptance by and social relations with nearby whites. As long as the "color line" holds fast for even those who are three-fourths or more white, one can expect such results as are presented above.

These results are in accordance with most of the more recent work on Indians which stresses the importance of environment over heredity and throws some doubt upon the existence of true biological differences, except in physical characteristics, between races. In any case it seems reasonable to suppose that the observable differences in mental, emotional and social traits between Indians and whites are reflections of differences in environment (even though there may be some innate factors as well), especially in the case of such matters as personality, interests, and emotional attitude which seem to be, in any person of any race, largely products of conditioning.

MENTAL ABILITY OF THE NATIVE RACES OF ALASKA

WALTER CROSBY EELLS
Stanford University

A. INTRODUCTORY

As an important part of a general survey of the system of education maintained by the United States Government for the natives of Alaska,[1] an extensive testing program was carried out in the local schools. This included two tests of mental ability, one of mechanical ability, three of physical ability, three of musical ability, and a group covering actual achievement in thirteen different phases of the school curriculum. The tests in mechanical, physical, and musical abilities and the school achievement tests will be discussed in two subsequent papers. This report will be limited to a summary of some of the more significant findings regarding the mental abilities of over one thousand children of the Eskimo, Aleut, and Indian races—the three native races which make up over half the population of the vast but

[1] H. Dewey Anderson and Walter Crosby Eells, *Federal Education in Alaska: A Study of the Administrative, Curricular, Economic, and Sociological Aspects of the Work of the United States Government for the Eskimo, Aleuts, and Indians of Alaska, 1930-32.* Made under the auspices of the School of Education of Stanford University, at the request of the United States Office of Education, supported by a grant from the Carnegie Corporation of New York. Mss. of 1662 pages in library of Stanford University, and of United States Office of Indian Affairs, Washington, D. C. Arrangements for publication in book form not yet completed.

Dr. Anderson was Director of the Survey. The writer was Associate Director and especially responsible for the testing program.

little known territory of Alaska. This is believed to be the most extensive and detailed test program yet carried out on a group of primitive peoples. While undertaken primarily to secure adequate information upon which to base recommendations for modifications in the curriculum of the educational system of the Territory, the results should be of considerable interest in themselves to students of racial psychology.

Collecting Data.—For the past forty years the education of the natives of Alaska has been carried on under the supervision of the Department of Interior, formerly through the Bureau of Education, more recently through the Office of Indian Affairs, but prior to the present investigation no comprehensive study had been made of the schools, approximately one hundred in number, widely scattered over this broad territory.

Slightly over half the total population of 59,278 (1930) is composed of the three native races—Eskimo, Aleut and Indian. The Indians of Southeastern Alaska are not essentially different from many of the Indians of the States concerning whose abilities and characteristics some studies have been made. Little, however, has been known definitely regarding the Eskimo and the Aleuts of the Far North who make up two-thirds of the native population of Alaska. This study was designed primarily to secure information regarding these two less known races.

Members of the field staff visited and studied directly in the field over two-thirds of the Eskimo and over three-fourths of the Aleuts for whom the government is conducting educational work. Approximately a fourth of the Indians were also studied in order to secure data for comparative purposes. An average of five days was spent at each school.

Almost twenty thousand miles were traveled by Doctor Anderson and Mrs. Anderson, a trained Binet tester, in 51 weeks of travel and study in Alaska and eight thou-

sand miles by the writer in six weeks in securing the data for the study. Unusual difficulties and dangers of travel had to be met in order to visit the widely scattered, almost inaccessible native schools while they were in session. Many weeks of weary travel by dog team in the hazards of an Alaskan winter were necessary. Other methods of transportation included railroad, steamship, coast guard cutter, cannery launches, gasoline boats, native craft, and aeroplane.

Tests were given to native children in 46 schools. An average of 13 measures was secured for each of the 634 Eskimo children tested; of 9 for each of the 195 Aleuts; and of 15 for each of the 255 Indians. These numbers are not large, judging by test results available in the States. In Alaska, however, the schools are small and scattered. The total number tested in Aleut and Eskimo schools was 60 per cent of the average attendance above the first grade. All available children were tested, as far as possible.

Limitations.—These thousand Alaskan native children do not form a homogeneous group, or anything approximating it, in background, experience, or training. They are composed of three quite distinct races. Some are pure blooded natives; others involve varying degrees of mixture of white blood. Some have been in close and frequent contact with white civilization; others are living under almost primitive conditions. Some have attended native schools only; others have received part or all of their education in territorial or in missionary schools. Some have been educated only in village day schools; others have been away from home for years in twenty-four-hour boarding and industrial schools. In certain parts of the Territory the native schools are in session, on an average, two months longer per year than in others, and in individual schools the differences are still greater.

Many of the tests assume a common racial and educational background and experience. This background in many respects is very different for the natives of the far North as compared with the average child in the United States. Not only educational opportunity but cultural patterns of thought are widely variant. There is a distinct linguistic handicap. The Eskimo children are taught in school in English—a language entirely different from their native tongue. In all of the village schools they use this language, at most, five or six hours a day during a part of the year only. The rest of the time they commonly use their native tongue. Thus they are taught and tested in what to them is a foreign language. An effort was made to offset this difficulty by a careful choice of tests in which the linguistic element was eliminated entirely, reduced to a minimum, or modified to suit local conditions and background. This, however, could not be accomplished completely.

Caution must be observed, too, in many of the comparisons between Eskimo, Aleuts, and Indians due to the varying proportion of children in "twenty-four-hour" schools. It is quite evident that pupils in industrial or missionary boarding schools have very different social, educational, and linguistic opportunities from those in the small village schools. This undoubtedly affects the showing they made on some of the tests. It happens that less than one-fourth (24%) of the Eskimo who were tested were in the industrial and boarding schools, but over half of the Aleuts (56%) and over half of the Indians (57%) were in these institutions. Comparisons between Indians and Aleuts, therefore, should be valid in situations in which comparisons of either of these two races with the Eskimo may be somewhat unfair to the latter.

In spite of these handicaps and limitations and others not mentioned here it is believed that there has been assembled a mass of data regarding a wide range of

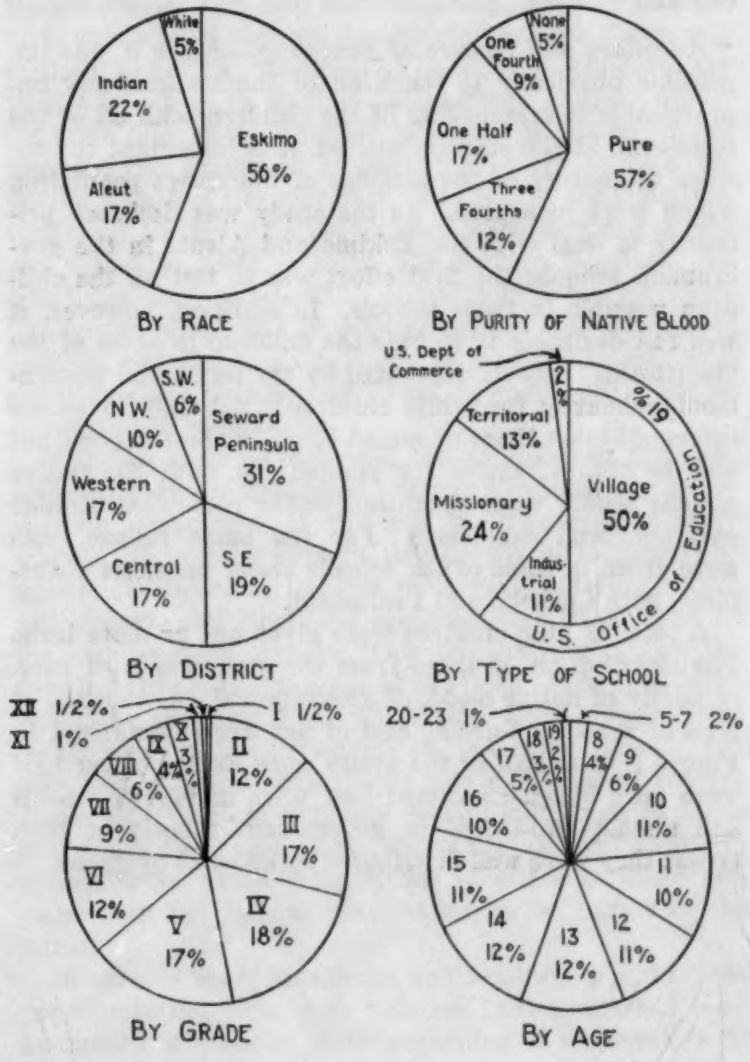
abilities and scholastic achievements of both Eskimo and Aleuts, far more comprehensive and reliable than has previously been known, which have real scientific value and significance when interpreted with due statistical caution.

Adequacy and Nature of Sampling.—Since it was impossible physically to reach all of the stations and impracticable to measure all of the children with all of the tests even at the stations visited, it is important to consider the nature of the samples of the entire population which were measured. As the study was designed primarily to deal with the Eskimo and Aleuts in the government schools, the first effort was to test all the children possible in these schools. In addition, however, it was felt desirable to include the children in some of the "territorial" schools (operated by the territorial government primarily for white children), not only to secure information on those of mixed blood in these schools but also to obtain results for comparison with the native schools which were operating under somewhat similar environmental conditions. For the same reason tests were given in some of the schools under missionary auspices, both Catholic and Protestant.

A total of 1138 children were given one or more tests. The distribution of these from the standpoints of race, of purity of native blood, of geographical distribution, of type of school, of grade, and of age are summarized in Figure I. Over half of the group were Eskimo; over half were pure blooded natives; half were in village schools and almost two-thirds in government schools of both types; they were well distributed by age and by grade.

FIGURE I

Classifications of Alaskan School Children Who Were Given a Group of Tests of Ability and Achievement (N = 1138)



On the whole it is felt that the measurement of children above the first grade in the government schools for the Eskimo and Aleuts is reasonably representative and adequate, and that satisfactory samples were secured of other races and types of schools. Certainly no effort was made to select children on any other basis than their availability in the villages and schools which it was possible to visit. It is doubtful whether complete measurement of the entire school population, even had it been possible, would yield results materially different from those reported.

Procedure.—With one or two minor exceptions all of the tests were given by members of the staff, thus assuring uniformity of procedure and fairness of administration.[2] The first tests were given at Nome in October, 1930; the last in Sitka in August, 1931. All group tests were sent to Stanford University where they were scored and checked by trained scorers under the direct supervision of the writer. Analyses of the results were all made by him.

B. Mental Ability—By Stanford-Binet Scale

Information regarding the mental ability of children is of fundamental importance in studying the actual curriculum in the Alaskan native schools and the success of the children in it. The practical school administrator is interested in the question of the ability of children to do school work under given conditions. If the tests will give satisfactory measures of such ability they are exceed-

[2] The various group tests were given by the writer at Nome and White Mountain; in all other places by Dr. Anderson. All individual Binet tests were given by Mrs. Anderson who was specially trained for this work prior to going to Alaska. She usually spent all of her time in individual testing with the Stanford-Binet at each station visited. In three small schools group tests were given by teachers experienced in testing or specially trained by one of the staff.

ingly useful, regardless of whether they measure traits which are innate or acquired or a combination of the two. If in addition they can throw light upon unsettled fundamental problems of essential differences in racial psychology, they will prove all the more useful and significant.

The measure upon which the greatest stress was placed was the Stanford Revision of the Binet-Simon Scale. Its proved high reliability, careful validation, wide range of qualitatively different test elements, and the close rapport between tester and subject made it the most significant single instrument available for use in the Alaskan study. Its chief limitation was the strong linguistic element involved and the numerous questions dealing with concepts and terms not in the common experience of Eskimo children. Certain minor modifications were made to adapt it better to the understanding and experience of native children.[3] It was felt that these tended to make it a fairer measuring instrument for the desired purpose under the conditions obtaining in Alaska.

Summary of Results.—Table I exhibits a general summary of the distribution of IQ's for the 566 children among the three native races who were measured with the Stanford-Binet Scale. The most comprehensive testing was done among the Eskimo who constituted over two-thirds of the entire group.[4] The distribution of IQ's

[3] For examples the fables were omitted as containing too many abstract or foreign words. The phrase, "the road from my house to the city" was changed to "the path from my house to the beach" since roads and cities are unknown to most Eskimo children. Similarly "engineer," "cars," and "train" were changed to "captain," "fish," and "boat;" "police" was changed to "marshall," the only term used for such an officer of the law; "snake" was changed to "muskrat;" and "cow" and "sparrow" to "reindeer" and "duck."

[4] Travel conditions made it impossible to secure as comprehensive or representative a sampling of the Aleuts as of the Eskimo. Most of those tested were in industrial or missionary schools. Only a moderate sampling of Indian children was attempted, a majority

for the Eskimo children is shown graphically in ten point intervals in Fig. II.

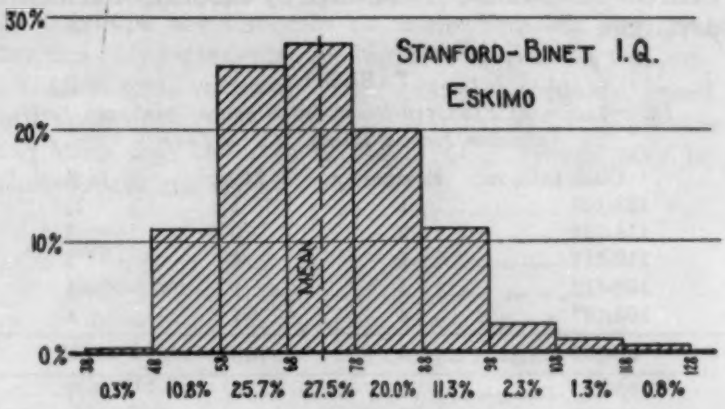


FIGURE II

Distribution of Intelligence Quotients of 389 Eskimo Children as Determined by the Stanford-Binet Scale

The mean IQ for the Indians and Aleuts of 79 and 80 with relatively small populations of each, show no significant differences from each other; the Eskimo with an average IQ of only 74, however, are significantly lower than the other two groups. Whether this difference is due to native racial differences in ability, to poorer cultural opportunities, or to sampling more truly representative of the entire population, the figures themselves do not indicate. A variety of reasons, however, leads to the conviction that the latter is probably the dominant factor and that if the sampling had been equally exten-

of them also being in full-time boarding schools. The Eskimo tested were more nearly representative of the general population, since they were drawn from many of the village schools as well as from the boarding schools. At each station all available pupils of test age were usually taken. In two or three cases where time did not permit such complete testing, one child was taken from each family.

sive and representative for the other races there would have been little if any significant differences in the general average. The Eskimo are somewhat less variable than the other races, as indicated by the smaller standard deviation.

TABLE I

Distribution of Stanford-Binet Intelligence Quotients for Alaskan Native Races. (N = 566)

Class Interval	Eskimo	Aleut	Indian
123-127	1	1	--
118-122	2	0	1
113-117	1	1	1
108-112	4	1	1
103-107	4	1	1
98-102	5	3	2
93- 97	21	8	7
88- 92	23	13	4
83- 87	35	15	15
78- 82	43	16	11
73- 77	58	9	9
68- 72	49	8	13
63- 67	49	6	12
58- 62	51	5	6
53- 57	34	6	
48- 52	8	1	
43- 47	0		
38- 42	1		
Number	389	94	83
Mean	73.67	80.27	78.88
Standard Deviation	12.78	13.94	14.92
P. E. of Mean	0.44	0.97	1.11

Comparisons with Other Races.—Garth, who has made extensive studies in racial psychology, has summarized the principal experimental and statistical studies which have been made of mental differences in races.[5]

[5] T. R. Garth, *Race Psychology: A Study of Racial Mental Differences*, McGraw-Hill, New York, 1931. Chapters IV, V.

Of over a hundred such studies which he lists, however, only a few used the individual Binet tests and in most of these cases the populations were relatively small. Conclusions regarding essential racial differences based on small groups are likely to be misleading—at any rate they can only be accepted as suggestive. After a full consideration and evaluation of all available evidence, based both upon group and individual tests, Garth reaches the conclusion that the IQ's of these racial groups may be taken approximately as follows:

Whites	100
Japanese	99
Chinese	99
Northern Negroes	85
Hawaiians	84
Mexicans	78
Southern Negroes	75
American Indians (full blood)	70

If these results may be accepted as reliable, and no better ones at present are obtainable, then the Eskimo as measured by the Stanford-Binet Scale may be ranked as distinctly superior in mental ability to the American Indian and approximately equal to the Southern Negro. When it is remembered that they have not had the opportunities for contact with white civilization which have characterized almost all of the other groups tested in the United States proper it is likely that their real ability is somewhat superior. The evidence to be presented in the next section tends to substantiate this judgment with the indication that the true average IQ is perhaps as much as 15 points higher than that reported above.

Sex Differences.—A summary of the distribution of IQ's as determined by the Stanford-Binet Scale is shown in Table II for boys and girls separately.

TABLE II

Summary of Stanford-Binet Intelligence Quotients for Alaskan Native Races by Sex. (N=566.)

Group	N	Mean	Sigma	P. E. of Mean	Difference of Means	P. E. of Difference
Eskimo						
Boys	154	75.4	13.1	0.7		
Girls	235	72.6	14.1	0.6	+2.8	0.9
Aleuts						
Boys	37	83.1	14.1	1.6		
Girls	57	78.4	13.6	1.2	+4.7	2.0
Indians						
Boys	43	79.1	11.5	1.2		
Girls	40	78.9	13.3	1.4	+0.2	1.8

In each race the mean IQ for the boys is higher than that for the girls but only in the case of the Eskimo is this difference, when taken in connection with the number of pupils measured, great enough to be statistically significant.

Superior Pupils.—In a normal white population in the United States it may be expected that 50 per cent of the pupils will have IQ's exceeding 100. Garth shows that only 7.5 per cent of Negroes and 8.9 per cent of Mexicans tested had IQ's greater than 100. Corresponding figures for the Alaskan native races as measured by the Stanford-Binet IQ are as follows:

Eskimo	4.1%
Aleuts	5.3%
Indians	7.2%

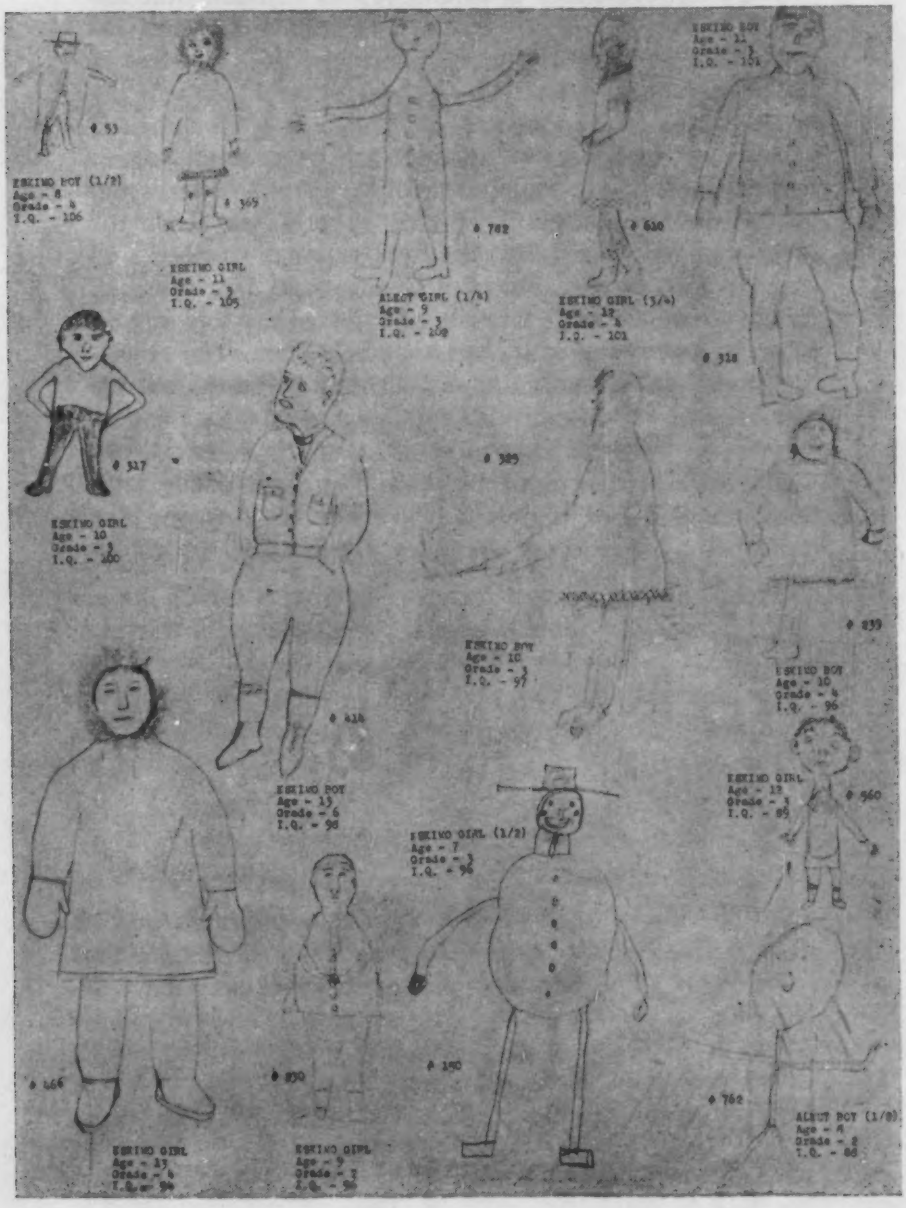
Of the 27 children, however, with IQ's exceeding 100, only four were pure blooded natives, three Eskimo and one Indian, and only one of the four had an IQ above 102. One pure blooded Eskimo girl, age 9 years 10 months, in the third grade in a village school, tested 118. A half-breed Eskimo girl and a quarter-breed Aleut boy each tested 127, the highest found in the entire study.

C. Mental Ability—By Goodenough Scale

The Stanford-Binet is considered of fundamental importance in the determination of mental ability but because of the linguistic elements involved in it, because of its partial dependence upon school experience, and because of the very different cultural and environmental background of the races studied in this investigation, it was thought wise to supplement it by a simple non-verbal group intelligence test. For this purpose the Scale for Measurement of Intelligence by Drawings devised and standardized by Dr. Florence Goodenough was selected. In the use of her scale a group of children are told simply to make a drawing of a man—a whole man. The resulting drawings are scored upon over fifty features, artistic merit as such receiving only minor consideration. A few sample drawings are shown in Figure III. In each case the race, sex, degree of blood purity, age, grade, and Goodenough IQ of the child who made the drawing are indicated. All children are pure blooded natives unless otherwise indicated.

FIGURE III

Sample Drawings by Eskimo and Aleut Children with Moderate Intelligence Quotients as Judged by the Goodenough Scale



Summary of Results.—Table III gives the distribution of the IQ's of the 527 native children upon whom this test was used. The mean scores for the three races are

TABLE III

Distribution of Goodenough IQ's for Alaskan Native Races (N=527)

Class Interval	Eskimo	Aleut	Indian
153-157	1		
148-152	1		
143-147	0		
138-142	1		
133-137	2		1
128-132	2	1	0
123-127	1	1	1
118-122	4	1	1
113-117	10	7	1
108-112	11	9	2
103-107	19	7	1
98-102	45	20	10
93- 97	68	17	11
88- 92	48	8	9
83- 87	42	12	7
78- 82	39	4	6
73- 77	16	8	2
68- 72	26	3	5
63- 67	14	2	0
58- 62	5	2	1
53- 57	4	1	
48- 52	4	2	
43- 47	1	2	
Number	364	105	58
Mean	89.56	93.29	91.55
Standard Deviation	15.57	15.59	13.74
P. E. of Mean	0.55	1.03	1.22

found in the same order as for the Stanford-Binet, but all are somewhat higher. Differences are as follows:

	Stanford-Binet	Goodenough	Difference
Eskimo	73.7	89.6	15.9
Aleuts	80.3	93.3	13.0
Indians	79.0	91.6	12.6

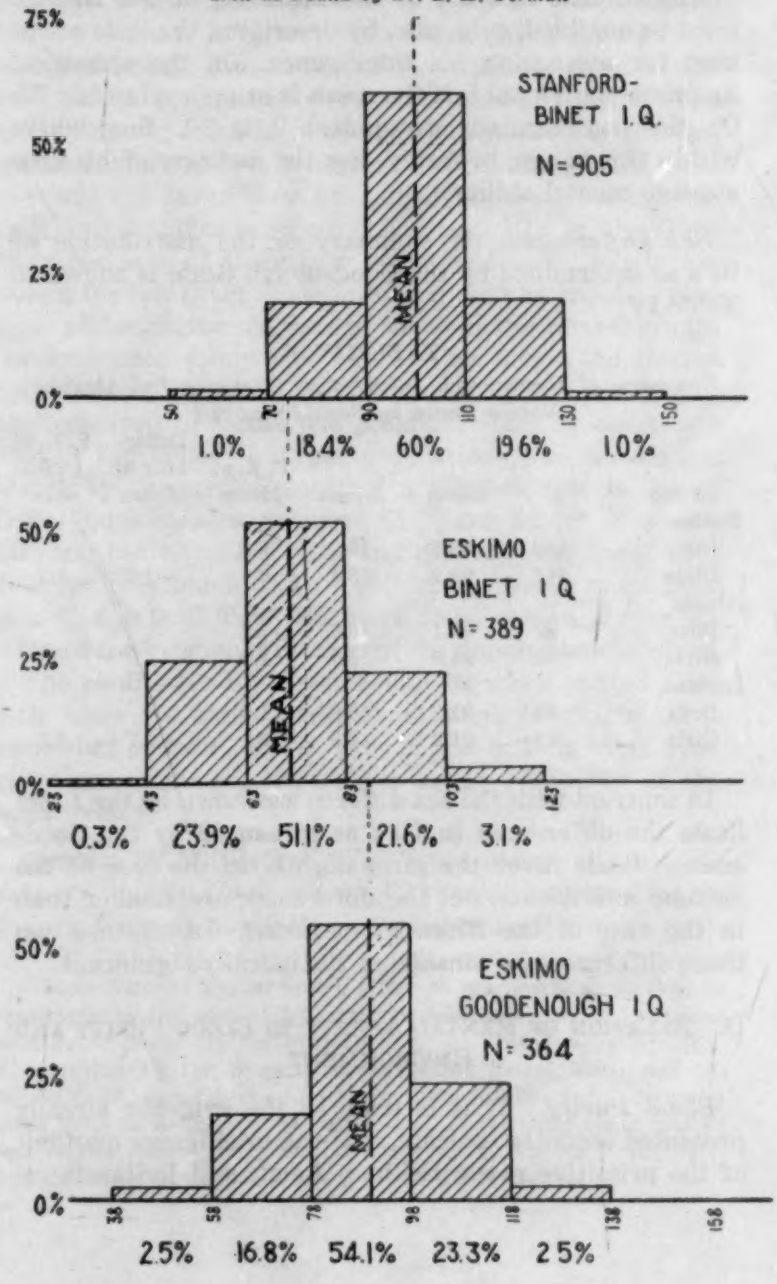
These comparisons suggest that possibly as much as 15 points of IQ may be due to the language factor involved in the more reliable individual Stanford-Binet. There is no definite evidence, however, that all of the difference is due to the language factor alone. Variations of cultural background and school experience doubtless have some influence as well. In many cases, too, represented in Tables I and III the same individuals were not tested by both scales. There were, however, 328 cases in which the same Eskimo pupils were measured by both scales. For this group the mean IQ by the Stanford-Binet was 74.63, while by the Goodenough it was 90.26, making a difference of 15.63, practically the same as for the two Eskimo groups already compared.

Nature of Distribution. Further light upon the comparable validity of the two tests is given by comparing the nature of the distribution of Eskimo, the largest racial group, on each of the two tests with the distribution of IQ's given by Terman in the original standardization of the Stanford-Binet. This is done in Figure IV where the grouping is in class intervals of 20 points of IQ as in Terman's original report.[6] These diagrams show strikingly the essential similarity in *form* of the three distributions, but an equally striking difference in their *position.*

[6] L. M. Terman, et al., *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence* (Genetic Psychology Monograph No. 18). Warwick and York, Baltimore, 1917, page 42.

FIGURE IV

Comparison of the Distribution of Intelligence Quotients of Eskimo Children as Measured by the Stanford-Binet and Goodenough Scales, with the Standard Distribution of Intelligence Quotients of White Children as Found by Dr. L. M. Terman



Any definite location of the mean IQ of the Eskimo must be qualified, evidently, by describing the instrument used for measuring his intelligence. On the somewhat linguistic individual test the mean is at approximately 75. On the non-linguistic group test it is 90. Somewhere within this range, probably, lies the measure of his true average mental ability.

Sex Differences. A summary of the distribution of IQ's as determined by the Goodenough Scale is shown in Table IV.

TABLE IV

Summary of Goodenough Intelligence Quotients for Alaskan Native Races by Sex. (N=527.)

Group	N	Mean	Sigma	P. E. of Mean	Difference of Means	P. E. of Difference
Eskimo						
Boys	152	88.7	15.8	0.9		
Girls	212	90.2	15.4	0.7	—1.5	1.1
Aleuts						
Boys	42	92.1	16.6	1.7		
Girls	63	94.0	15.0	1.3	—1.9	2.1
Indians						
Boys	29	91.9	10.3	1.3		
Girls	29	91.2	16.5	2.1	+0.7	2.5

In contrast with the sex differences shown by the Binet Scale the differences in IQ's as measured by the Goodenough Scale favor the girls slightly in the case of the Eskimo and Aleuts, but the differences are smaller than in the case of the Binet comparisons. In no case can these differences be considered statistically significant.

D. Relation of Mental Ability to Blood Purity and Environment

Blood Purity. If it is true, as the evidence already presented seems to indicate, that the intelligence quotient of the primitive races, Eskimo, Aleut, and Indian is es-

sentially lower than for whites, then its value should tend to increase progressively with additional admixture of white blood. Two-thirds of the children tested as far as could be determined were pure blooded natives. The remaining one-third have been classified as three-fourths, one-half, and one-fourth native blood. Although the numbers are not large they are sufficient to indicate a distinct and significant trend. The mean IQ's of the three races for each of these four degrees of racial purity are shown for the Binet Scale in Figure V. The trend is very clear although the number of cases in the mixed groups for each race (shown in figures just below the plotted points) is rather small. If the figures for the three races are combined to secure the greater reliability characteristic of the mean of larger populations, the results are shown by the circled dots. If a straight line is drawn from "pure-blooded natives" to "pure white" it passes through the circled dot for half-breeds, and the circled dots for one-fourth and three-fourths breeds vary from this "theoretical line" no more than can easily be accounted for by chance in view of the populations involved.

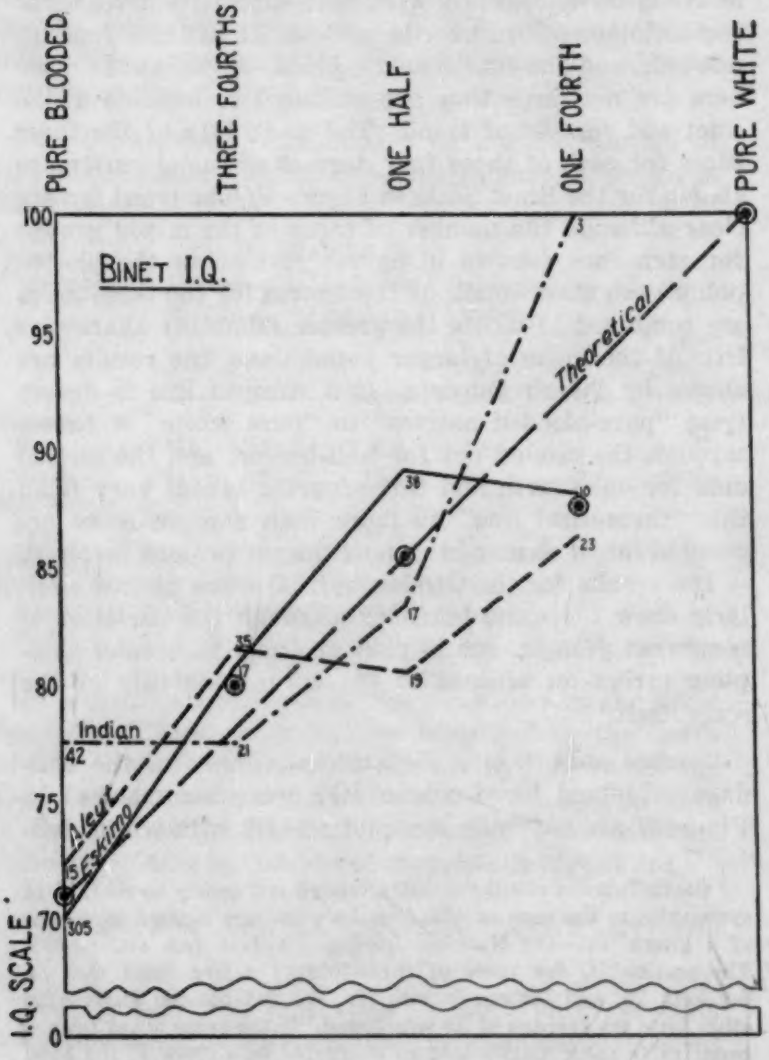
The results for the Goodenough IQ when plotted similarly show the same tendency although the variation is somewhat greater, due in part probably to greater sampling errors on account of the lower reliability of the scale itself.[7]

Contact with White Civilization. Many of the children of mixed blood whose IQ's were summarized in Figure V are naturally in closer contact with white civil-

[7] Garth found a similar trend, although not nearly so marked or systematic in the case of 527 American Indians studied by means of a group test—the National Intelligence Test (*op. cit.*, p. 36). The median IQ for those of three-fourths native blood was 74, for half, 75, and for one-fourth, 78. For full-blooded Indians by other tests an average of 70 was found. If the same trend toward progressive racial purity held as suggested by Figure V, the half-breeds should have had a mean IQ of 85 instead of 75.

FIGURE V

Relationship of Intelligence Quotient, as Determined by the Stanford-Binet Scale, to Degree of Racial Purity for Eskimo, Aleut, and Indian Children



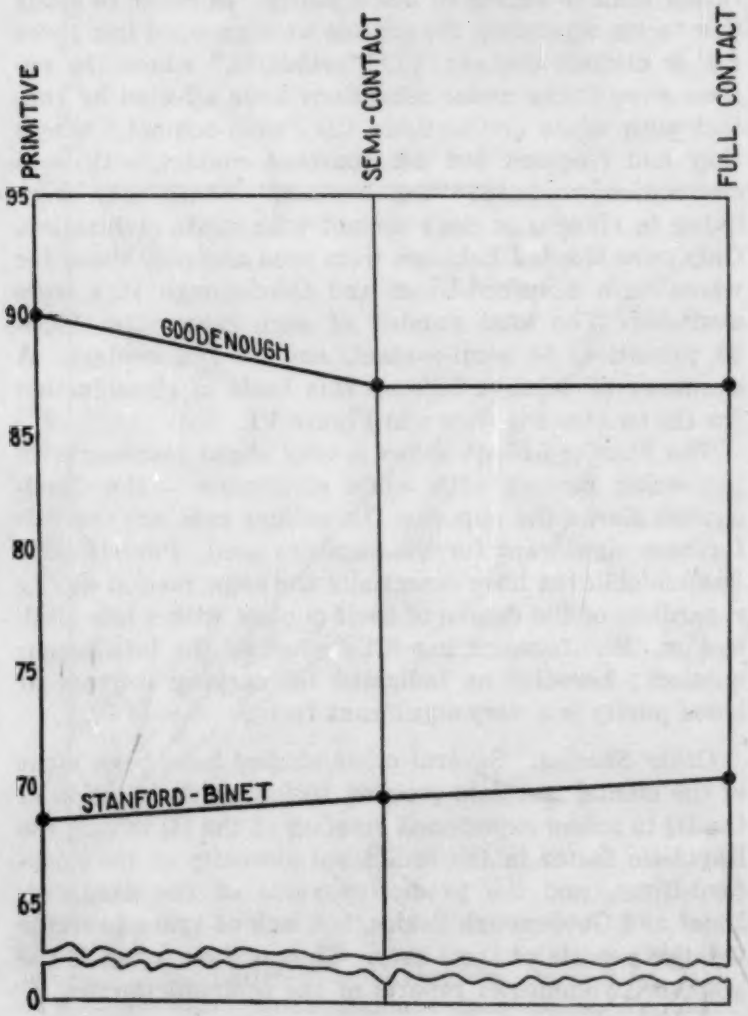
ization than the pure-blooded natives. Some of them were attending territorial schools. It is possible, therefore, that much of the increasing superiority shown in Figure V is due to closer contact with white civilization rather than to degree of blood purity. In order to study this factor separately the schools were grouped into three rather distinct classes: (1) "primitive," where the natives were living under conditions little affected by contact with white civilization; (2) "semi-contact," where they had frequent but not constant contact with such civilization; and (3) "full contact," where they were living in villages in close contact with white civilization. Only pure-blooded Eskimos were used and only those for whom both Stanford-Binet and Goodenough IQ's were available. The total number of such cases was 199,—85 primitive, 64 semi-contact, and 50 full contact. A summary of relative IQ's on this basis of classification for the two tests is shown in Figure VI.

The Stanford-Binet shows a very slight increase with increasing contact with white civilization — the Goodenough shows the opposite. In neither case are the differences significant for the numbers used. Pure-blooded Eskimo children have essentially the same mental ability regardless of the degree of their contact with white civilization. Environment has little effect on the intelligence quotient; heredity as indicated by varying degrees of blood purity is a very significant factor.

Other Studies. Several other studies have been made of the mental test data secured, including the relation of the IQ to school experience, relation of the IQ to age, the linguistic factor in the individual elements of the Stanford-Binet, and the predictive value of the Stanford-Binet and Goodenough Scales, but lack of space prevents detailed reports of them here. They will be found in the published volumes of reports of the complete survey.

FIGURE VI

Relationship of Intelligence Quotient, as Determined by the Stanford-Binet and Goodenough Scales, to Degree of Contact with White Civilization for Pure Blooded Eskimo Children (N = 199)



PREDICTING COLLEGE MARKS AND TEACHING SUCCESS FOR STUDENTS IN A TEACHERS COLLEGE*

RALPH A. FRITZ
Kansas State Teachers College

PURPOSE

This study endeavored to determine which of several criteria would likely be most helpful in predicting (1) college success and (2) probable teaching success for students majoring in education at Kansas State Teachers College, Pittsburg. One hundred women students were studied. The factors investigated were intelligence, age, the curriculum followed, the year of graduation from high school, the particular high school attended, race, teaching experience, occupation of parents, and high school record.

MEASURES USED

The measure of each student's college success was the average of her college marks up to and including the semester ending in June, 1930, as shown on the records of the college registrar. The number of semester hours of credit per student ranged from 16 to 120 with an average of 47. Each semester hour of credit with a mark of "A" was given a weight of 5, and credits with marks of "B," "C," "D," and "F" were weighted 4, 3, 2, and 1, respectively. An average of "C" would therefore be represented by 3.0.

* An abstract of a thesis by James W. Evans with added analyses and observations by the author of this article.

The measure of each student's probable teaching success was her score on "Aptitude Tests for Elementary and High School Teachers, Set 1," by J. E. Bathurst, F. B. Knight, G. M. Ruch, and Fred Telford.[1]

The measure of each student's intelligence was her score on the 1929 edition of the American Council Psychological Examination.[2]

The 100 students of this study comprised the total women students taking work in the Horace Mann Training School during the second semester of 1929-1930 with the exception of ten who were absent when one or the other of the tests was given. They were Education majors enrolled in one of the following curricula: (1) Kindergarten-Primary; (2) Intermediate; or (3) Three-year State Certificate.

Results

Table I contains much of the data on which the calculations of this study are based. It is to be read in this manner: Student number 1 was 18 years old; she made a gross score on the Psychological Examination of 119; her gross score on the Aptitude Test was 104; and her average college mark is expressed by 3.3.

[1] Bathurst, J. E., Knight, F. B., Ruch, G. M., and Telford, Fred, *Aptitude Tests for Elementary and High School Teachers, Set 1*, Washington: Bureau of Public Personnel Administration, 1927.

[2] Thurstone, L. L. and Thurstone, Thelma Gwinn, *Psychological Examination for High School Graduates and College Freshmen*, Washington: American Council on Education, 1929 edition.

TABLE I

Age, Gross Psychological Examination Score, Gross Aptitude Test Score, and Average College Mark for Each of the 100 Students.

Student No.	Age	Gross Psy. Exam.	Gross Apt. Test	Average College Mark	Student No.	Age	Gross Psy. Exam.	Gross Apt. Test	Average College Mark
1	18	119	104	3.3	51	20	71	82	2.0
2	19	106	120	3.4	52	19	148	106	3.1
3	24	143	103	4.2	53	20	167	121	3.4
4	18	139	121	3.1	54	19	140	116	3.8
5	20	95	119	3.4	55	21	44	70	2.9
6	19	144	137	3.4	56	20	90	106	2.9
7	20	133	107	3.7	57	20	84	115	3.6
8	18	189	149	3.6	58	20	111	104	3.3
9	25	174	153	4.0	59	21	142	126	3.6
10	21	166	124	3.2	60	19	77	91	2.9
11	20	93	89	3.2	61	26	132	122	3.7
12	21	100	102	2.7	62	28	105	90	3.5
13	21	142	110	3.5	63	20	153	132	3.7
14	25	164	134	3.9	64	19	114	138	3.5
15	23	118	107	3.2	65	18	102	96	3.3
16	18	101	114	3.8	66	21	93	97	3.8
17	19	135	112	3.2	67	20	148	125	4.1
18	21	93	102	3.8	68	21	56	92	3.0
19	20	74	103	3.3	69	18	177	121	3.8
20	21	55	111	3.2	70	18	82	92	3.3
21	19	121	138	3.4	71	23	75	107	2.8
22	26	118	92	3.3	72	18	90	107	3.7
23	19	104	99	3.4	73	19	92	99	1.9
24	18	111	102	4.2	74	23	95	67	3.1
25	20	82	87	2.7	75	19	86	67	2.5
26	20	98	92	3.3	76	20	72	99	3.1
27	17	170	124	3.3	77	20	111	121	3.3
28	20	133	113	2.8	78	19	138	100	3.2
29	20	82	106	2.9	79	19	167	152	3.9
30	19	127	99	3.2	80	20	95	88	3.9
31	19	117	122	3.6	81	29	26	69	3.2
32	20	237	147	3.4	82	19	103	95	3.0
33	20	206	153	4.4	83	19	65	60	2.3
34	19	167	123	3.6	84	22	59	107	3.1
35	21	89	84	3.1	85	18	116	94	2.9
36	24	165	158	3.7	86	20	160	139	4.1
37	20	122	102	2.8	87	18	168	155	3.5
38	20	135	128	3.7	88	20	95	86	2.6
39	18	157	136	3.4	89	19	105	125	3.5
40	22	127	160	4.0	90	17	111	116	3.6
41	19	167	108	3.7	91	18	120	132	3.7
42	19	174	138	3.4	92	20	168	130	2.8
43	18	162	131	3.7	93	20	106	85	3.3
44	25	106	103	3.5	94	17	117	107	3.9
45	17	75	74	3.2	95	18	239	159	4.2
46	19	75	65	3.1	96	18	102	100	3.4
47	18	99	79	2.4	97	17	150	93	2.9
48	18	141	123	3.3	98	19	94	106	2.6
49	20	113	89	3.0	99	21	147	113	3.1
50	39	129	127	3.6	100	20	146	136	3.4

Prediction of College Success

Psychological Examination

None of the factors investigated show great promise for predicting college success; however, both the Psychological Examination and the Aptitude Test may be helpful. The relationship between the average college marks for the 100 students and their scores on the Psychological Examination is represented by a correlation coefficient (Pearsonian) of .527±.048. This is low. However, of the eight students ranking highest on the Psychological Examination, six made an average college mark of "B" and two of high "C;" of the seven ranking lowest in the examination, two made average marks of "D" and five of "C." Of the 100 students, 38 averaged "B," 58 averaged "C," and 4 averaged "D." There were no averages of "A" or "F."

Aptitude Test

The correlation between the average college marks and the Aptitude Test scores was .629±.041 for the 100 students. This means that the Aptitude Test predicts college marks better than does the Psychological Examination. It may also mean that the Aptitude Test tended to measure what these people had learned in their college classes in education since all had taken such courses before they wrote the test. Of the 20 students scoring highest on the Aptitude Test, 14 had an average mark of "B" and the other 6 of high "C;" of the 22 scoring lowest, only 1 had an average of "B," with 18 of "C," and 3 of "D."

Age

There was a tendency for the younger students to secure slightly higher scores on both the tests. However their college marks and also those of the older students were so near the average that age could not be used

for predictive purposes. Other elements also complicated the analysis of this factor.

Curriculum Studied

Fifty-seven students were enrolled in the Kindergarten-Primary curriculum, 34 in the Intermediate, and 9 in the Three-year State Certificate. The average college mark for each group was 3.4, 3.2, and 3.4, respectively. These differences are too small to be of any value for predictive purposes.

Year of Graduation and High School Attended

Neither the year a student graduated from high school nor the high school from which she came gave any hint as to the average college mark she earned. However, the 20 students who came from the Pittsburg High School were slightly above average in both college marks and intelligence.

Negroes

Nine of the students were colored. The average of their college marks was 2.9. Only 19 of the 100 students had average marks as low as this. These colored students made an average score on the Psychological Examination of 105.2; 41 of the 100 students had scores as low as this. These colored students made an average score of 110.8 on the Aptitude Test; 33 of the 100 students made scores as low as this. It appears that these colored students received lower college marks than their intelligence would warrant.

Teaching Experience

Twenty-five students had teaching experience other than supervised teaching. The amount of experience ranged from three months to 99 months, with an average of 25, and a median of 16. These experienced students averaged 2.5 years older than the total 100 students. Their average college mark was 3.5, as compared with

3.3 for the 100 students. Their intelligence and Aptitude scores were very slightly above average. It is probable that their slight superiority in scholarship is not attributable to greater intelligence so much as to better work habits and to a greater proportion of "A" and "B" marks a few years ago when these students took some of their college work.

Occupation of Parents

The parents of these 100 students were engaged in 30 different occupations. The parents of 23 of them were farmers. The average college mark of this group of 23 students and their scores on both the Psychological Examination and the Aptitude Test were well distributed throughout the ranges for the 100 students and afford no factors of value for predictive purposes.

High School Record

The 100 students came from 56 different high schools in five states. No satisfactory plan for equating their high school marks could be found and the records showing the pupils' success in high school had to be eliminated from this study.

Prediction of Probable Teaching Success

As previously stated, the criterion of teaching success used in this study is each student's score on the teachers' Aptitude Test. This test is supposed to measure the extent to which a student possesses the qualities of a successful teacher. It is objective and probably more valid for this purpose than the usual rating schemes.

Psychological Examination

If the Aptitude Test does measure teaching success, then the Psychological Examination is a fair predictor of teaching success. The correlation between these two tests for the 100 students was .732±.081. This compara-

tively high correlation might be taken as indicative that the Aptitude Test is really another intelligence test. Table II gives the correlations between the tests by sections.

TABLE II

Pearson Correlations Between Gross Scores on the Psychological Examination and Gross and Sectional Scores on the Teachers' Aptitude Test (N=100).

Section	No. of Items	Type of Items	Time in Min.	r	P.E.$_r$
1. Professional Judgment	28	5-response	15	.392	±.057
2. Theory & Practice of Teaching	85	True-false	20	.531	±.048
3. Reading Comprehension	25	Questions	25	.585	±.044
4. Social Information	24	4-response	8	.513	±.050
5. School & Class Management	55	True-false	15	.694	±.035
6. Professional Information	33	True-false	7	.426	±.055
Totals	250		90		
Gross Score				.732	±.031

College Marks

To what extent do college marks predict probable teaching success? The correlation between the college marks and the Aptitude Test scores was .629±.041 for the 100 students. This is not high; however it does show considerable relationship between college marks and teaching success—providing the Aptitude Tests constitute a valid measure of teaching success.

Other Factors

From a further analysis of the groups previously mentioned in our search for factors to predict college success, we find no significant relationship between probable

teaching success as measured by the Aptitude Test and the factors of: curriculum followed, year of graduation from high school, particular high school attended, the fact that the students were colored, that they had had teaching experience, or that their parents were farmers.

SUMMARY

The findings given here apply only to students in the one department of the Kansas State Teachers College of Pittsburg and cannot be considered applicable to all departments in that college, nor to all colleges.

Of the several factors investigated for predicting college success, only the two tests appeared to have any value. These were the Aptitude Tests for Elementary and High School Teachers, Set 1, and the American Council Psychological Examination, 1929 edition. The Aptitude Test gave a correlation of .629±.041 with college success, and the Psychological Examination a correlation of .527±.048. The best that can be said even for these two tests is that they may be helpful in predicting college marks but are by no means dependable predictors.

The measure of a student's probable teaching success was her score on the Aptitude Test. Of the several factors investigated for predicting probable teaching success, only two had any value. These were the Psychological Examination and average college marks. The Psychological Examination gave a correlation of .732±.031 with the Aptitude Test. One is inclined to wonder whether the Aptitude Test is not another intelligence test rather than a measure of teaching success or ability. The average college mark correlated .629±.041 with the Aptitude Test. If the Aptitude Test really measures ability to teach school, then it is evident that those students who earn the better marks in college are the better risks as teachers.

THE VALIDITY AND RELIABILITY OF TESTS OF MUSICAL TALENT

RALEIGH M. DRAKE
Wesleyan College

INTRODUCTION

The purpose of this study is to present the comparative validities and reliabilities of some twenty tests of musical talent. The meaning of the latter term is not taken to be based upon any particular definition but it will be assumed to exist where it can be observed by competent teachers who know their pupils well enough to accurately differentiate between their abilities, or who by the same observation have found that certain tests given to aural training classes discriminate between the musically gifted and ungifted. There seems to be no difficulty in reaching a fairly complete agreement in respect to musical talent in action, *i. e.*, for teachers and musical critics to agree upon the respective amount of talent possessed by different members of a certain group. But when an attempt is made to define objectively what they mean, there appears little in common. Thus Wundt (9) considered anyone musical if he were able to recognize and remember an octavo, a fifth, and a third, even if only for the time of the experiment! Stumpf (8) defined a musician by the ability to: (1) sing a given note struck on the piano, (2) judge the higher of two successive tones, (3) judge whether certain tones are present in chords, and (4) to determine the relative pleasantness of two chords. Crzellitzer (4) takes as the main item among several the ability to reproduce an easy melody at will, after a period of time, by singing, or by playing on an

instrument. Binet (2) requires only that a person have a normal voice and normal hearing! Kohler (6) takes it to be evidenced in the ability to recognize tone character; thus each note has a vowel characteristic so that one may speak of the C-ness or D-ness of these notes. Billroth (1) believes that one should have a talent for rhythm, pitch, intensity, and tone quality. His publication in 1896 seems to have been the inspiration for considerable amount of the work in the testing of musical talent that followed. Copp (3) states that musical ability is more common than ordinarily thought, for eighty per cent of children can be taught to sing middle C on demand. Matthay (7) states that it is the ability to learn with ease which depends upon the power of memory, which in turn depends upon attention, and this must further be combined with intellectual and emotional imagination. In this paper no definition will be formulated, but the teachers' location of talent in their pupils will be used as a criterian. When tests are found which correlate highly with these estimates, it will then be possible to determine the nature of the talent from the nature of the tests.

Description of the Experimental Groups

Royal Academy of Music Group A (R. A. M. Group A). This group consisted of the first, second, and third year aural training classes at the Royal Academy of Music, London, in 1929. All of the students had received several years of musical training before entering the Academy; their ages ranged from 18 to 25 years. Ninety per cent of the students were women. At the beginning of the testing, fall of 1929, there were 100 subjects, but at the end of the year when the final examination was given the largest number taking any one test was 53.

Royal Academy of Music Group B (R. A. M. Group B). This group consisted of the first and second year aural training classes at the Royal Academy during the

school year 1930-1931. There were 30 in this group of the same description as R. A. M. Group A.

Incorporated London Academy of Music Group A (I. L. A. M. Group A). This school is also known as the T. H. Yorke Trotter School. This group consisted of children from the ages of 9 to 16 years with about an equal number of boys and girls. The pupils were scholarship selections from the London County Council Schools and were chosen by an examination which has been described in a former article (5). Their musical training at the school varied from two months to three years and a few had received additional study up to eight years on some musical instrument. The training at the school is not of the usual sort but consists of what is broadly termed musical appreciation and ear training. They are taught to recognize and name the notes of the scale from hearing them, to tell when and into what key a melody changes, to sing answers to a given phrase, and to write short melodies from dictation. After the first year this work is usually supplemented by instrumental study.

Incorporated London Academy of Music Group B (I. L. A. M. Group B). The tests were given to this group in the school year of 1930-1931, a year later than the tests given to I. L. A. M. Group A. The subjects were substantially the same as in Group A, the only difference being that there were a few new students while others had left. The two groups are to all practical purposes synonymous, but their results are kept separate.

West Kensington Central School for Boys Group (W. K. C. S.). This group consisted of the first three forms (grades) in the West Kensington Central School for Boys, one of the London County Council Schools. The boys' ages varied from 11 to 15, with an average of 13 years. The majority of the students had received no

THE TESTS EMPLOYED

Test	Author	Reference
1. Musical Memory	Raleigh M. Drake	*J. Applied Psy.*, XVII, 136-147, 1933
2. Interval Discrimination	Raleigh M. Drake	*J. Applied Psy.*, XVII, 136-147, 1933
3. Retentivity	Raleigh M. Drake	*J. Applied Psy.*, XVII, 136-147, 1933
4. Intuition	R. M. Drake and S. Chapple	*J. Applied Psy.*, XVII, 136-147, 1933
5. Cadence	H. Lowery	*Brit. J. Psych.*, 1926, 17
6. Quality	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
7. Melodic Taste	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
8. Pitch	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
9. Tonal Memory	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
10. Tonal Movement	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
11. Intensity	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
12. Pitch Imagery	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
13. Rhythm Imagery	J. Kwalwasser and P. Dykema	RCA-Victor Co., Inc., Camden, N. J. (Phono. records)
14. Pitch	Carl Seashore	C. H. Stoelting Co. (Columbia records)
15. Intensity	Carl Seashore	C. H. Stoelting Co. (Columbia records)
16. Tonal Memory	Carl Seashore	C. H. Stoelting Co. (Columbia records)
17. Time	Carl Seashore	C. H. Stoelting Co. (Columbia records)
18. Consonance	Carl Seashore	C. H. Stoelting Co. (Columbia records)
19. Rhythm	Carl Seashore	C. H. Stoelting Co. (Columbia records)
20. Tonal Sequence	Max Schoen	*J. Comp. Psych.*, 1925, 5

musical training other than the usual school training. A few, however, had taken private lessons for from one to eight years and were members of the school band and glee club. There were 180 in this group.

VALIDITY

Nature of Criteria. In the I. L. A. M. groups the method of determining the validity of the tests was to have the principal, who knew all of the students well from his association with them as teacher, rank them according to their abilities. He was instructed to consider only the innate capacity of each, making no allowances for differences due to age or training. Proficiency on an instrument was not considered, but rather the expression in playing and the rapidity of learning were emphasized. He was directed to select the student in the group having the greatest talent for music and to give him a ranking of 1; then to select the student having the least talent for music and give him a ranking of 46 for Group A and 34 for Group B. He was then to select the student having the next best talent and give him a ranking of 2, and the third best and give him a ranking of 3, and so on until each member of the group had received some rank that represented his ability with respect to the other members of the group.

An entirely new ranking was secured for I. L. A. M. Group B than that used for I. L. A. M. Group A. It was done in exactly the same way and by the same person. As the two groups were almost identical, there were 38 of the same students in both groups, a ready means was secured to check the reliability of his estimates. The correlation between his first ranking (Group A) and his second ranking (Group B) was .684 P. E. $\pm$.057. There was an interval of eight months between the two rankings.

The criterion of musical talent for both groups at the

TABLE I

Validity Coefficients

	I.L.A.M. Raw Validity Coefficients		I.L.A.M. Net Validity Coefficients (age and training eliminated)		R.A.M. Raw Validity Coefficients	
	Group A	Group B	Group A	Group B	Group A	Group B
	r P.E.	r P.E.	r P.E.	r P.E.	r P.E.	r P.E.
1. Musical Memory (Drake)	.671±.054 N=46	.546±.082 N=34	.499±.075 N=46	.511±.080 N=34	.660±.076 N=26	
2. Interval Discrimination (Drake)	.585±.082 N=32	.421±.084 N=42	.406±.102 N=32	.351±.094 N=42	.770±.052 N=29	.409±.126 N=22
3. Retentivity (Drake)	.546±.084 N=36	.387±.076 N=55	.271±.099 N=36	.324±.082 N=55		.520±.096 N=24
4. Intuition (Chapple-Drake)	.361±.108 N=27	.355±.094 N=42	.245±.115 N=27	.304±.097 N=42		
5. Cadence (Lowery)	.482±.095 N=32	.273±.115 N=28	.389±.103 N=32	.211±.118 N=28		
6. Quality (Kwalwasser-Dykema)	.171±.102 N=35	.241±.085 N=54	—.096±.113 N=35	.153±.088 N=54		
7. Melodic Taste (Kwalwasser-Dykema)	—.132±.111 N=34	—.108±.090 N=55	—.131±.113 N=34	—.107±.090 N=55		—.089±.135 N=17
8. Pitch (Kwalwasser-Dykema)	—.239±.115 N=30	—.003±.095 N=50	—.101±.122 N=30	—.072±.100 N=50		
9. Tonal Memory (Kwalwasser-Dykema)	.417±.096 N=34	.255±.085 N=55	.230±.109 N=34	.168±.088 N=55		.503±.113 N=18
10. Tonal Movement (Kwalwasser-Dykema)	.155±.111 N=35	.238±.089 N=54	—.168±.110 N=35	.137±.094 N=54		—.027±.135 N=21
11. Intensity (Kwalwasser-Dykema)						.435±.127 N=15
12. Pitch Imagery (Kwalwasser-Dykema)						.806±.049 N=23
13. Rhythm Imagery (Kwalwasser-Dykema)						.158±.133 N=23
14. Pitch (Seashore)	.270±.115 N=28	.315±.097 N=40	.009±.123 N=28	.184±.104 N=40	.534±.097 N=25	
15. Intensity (Seashore)	.051±.134 N=26	.143±.103 N=38	—.251±.126 N=26	—.043±.107 N=38	.157±.090 N=53	
16. Tonal Memory (Seashore)	.334±.120 N=25	.418±.090 N=38	.231±.130 N=25	.376±.093 N=38	.114±.134 N=25	
17. Time (Seashore)	.149±.147 N=26	.383±.098 N=35	—.216±.130 N=26	.304±.104 N=35	.414±.127 N=20	
18. Rhythm (Seashore)	.084±.136 N=26	.367±.093 N=39	—.163±.132 N=26	.305±.097 N=39		
19. Consonance (Seashore)		—.030±.110 N=35				
20. Tonal Sequence (Schoen)	.243±.141 N=21	—.112±.113 N=15				

Royal Academy of Music was an examination given at the end of the school year. The instructor of the aural training classes, Ernest Reed, F. R. A. M., considers that the type of examination he gives is a good measure of musicality as evidenced in the ability to profit by training, and gives the same results in a more objective manner than direct estimates. The examinations for all the classes were of the following nature, graded in difficulty for each year:—

(1) Writing of chords and melodies from hearing them played on the piano; advanced classes were required to write counterpoint and four-part harmonies.

(2) Memorizing a melody in five minutes from notation on the blackboard, and then transposing into a given key either by writing it or playing it on the piano.

RESULTS

Table I shows the results for all of the tests. Only a few of the tests need be discussed. First it should be noted that for a correlation to be considered significant it should be at least three times its probable error with extraneous factors eliminated, such as age and training in this case, or at least five times the probable error when age and training are not controlled. The tests in Table II would be considered significant according to this statistical requirement.

Of the eight tests in Table II, the first two are the only ones which give significant coefficients in more than one group and in every group to which they are given (Pitch Imagery was given to only one group). Retentivity gave four out of five significant coefficients, Cadence gave two out of four, Intuition gave one out of three, Pitch (Seashore) and Tonal Memory (Seashore) each gave one out of five.

TABLE II

Test	Ratio of r (validity coefficient) to its P.E.					
	Age and training eliminated		Age and training *not eliminated*			
	I.L.A.M.		I.L.A.M.		R.A.M.	
	Group A	Group B	Group A	Group B	Group A	Group B
Musical Memory	6.6	6.3	12.4	6.6	8.7	
Interval Discrimination	4	3.7	7.1	5	14.6	(3.2)
Retentivity	(2.8)	4	6.5	5.7		5.4
Intuition	(2)	3	(3)	(3.7)		
Cadence	3.7	(2)	5	(2.3)		
Pitch Imagery						16.4
Pitch (Seashore)	(—13)	(1.7)	(2.3)	(3.2)	5.5	
Tonal Memory (Seashore)	(1.8)	4	(2.7)	(4.6)	(1)	

Note—The ratios in parenthesis are not significant.

The marked ratio shown by Interval Discrimination in R. A. M. Group A is largely due to the influence of specific training in interval recognition, the range of talent, and the selection in the higher classes of the gifted. All combined spuriously increase the correlation considerably. The true relationship is more likely to be found in I. L. A. M. Group A or Group B where age and training have been eliminated. In R. A. M. Group B the range of talent and selection are not so great and no such high correlation was found.

Pitch Imagery produces a very marked coefficient mainly because the material in this test was almost identical to that given in the examination. The test calls for a comparison of pitches heard on a phonograph with the printed notation, which is exactly the same task required in the examination at the R. A. M. The test definitely depends upon specific training in recognizing musical notation from merely hearing the pitches and could never be used as a test of talent in a group not having been thus trained. If this fault could be eliminated this test might prove to be quite useful.

Musical Memory is the only test from this group that consistently produces a high and significant validity coefficient. Other tests undoubtedly have suggestive value but in any pool of tests the weight given them cannot be nearly so much.

RELIABILITY

Method of Determining Reliability. The odd-even or split-half method was used for all of the tests. All responses were recorded in columns of ten, with the exception of Musical Memory which has a special form described elsewhere. The exact way in which each test was split into halves is described as follows:—

1. New Tests by Drake.
   a. Musical Memory was divided into two halves, one of which contained trials 1-4, 9-13, and 17-19 while the other consisted of trials 5-8, 13-16, and 21-24.

b. Interval Discrimination. The first 40 trials were correlated with the last 40 which in this case is the test-retest method with the order of the trials reversed. This gave the same result as correlating the upper five rows against the lower five.
c. Retentivity. The first 10 trials against the last 10 trials.

2. Test of Cadence by Lowery. The first 25 trials against the last 25.
3. The Seashore Tests. The tests of Pitch, Intensity, Tonal Memory, and Time were divided so that the upper five rows constituted one-half, and the lower five rows the other half. The remaining two tests, Rhythm and Consonance, were divided by taking the first 25 trials as one half and the last 25 as the other, or what amounts to the same thing, the A side of the record was correlated with the B side.
4. Kwalwasser-Dykema Tests. All of these tests except Melodic Taste were divided into two halves by the upper five rows and the lower five rows. Some tests, such as Rhythm, have only 25 trials so that the upper five rows contain five half-columns, while the lower five rows contain four half-columns. Melodic taste contains only 10 trials which the directions say to repeat to make the test longer; by taking the first 10 trials against their repetition one has the test-retest method.
5. Tonal Sequence by Schoen. This test consists of four trials, with four graded judgments to be made for each, making 16 scores for the total. Each trial of four answers was recorded in a separate column. Reliability was measured by correlating the two upper rows with the two lower.

RESULTS

Table III presents a summary of all the results. Here the most reliable group to use for a critical study of reliability would be the W. K. C. S. Group for it is an unselected group which will most frequently be found in actual practice. According to the usual statistical procedure a test to be of any practical use in actual prognosis should have a reliability of at least .85. Taking the corrected or raised reliability coefficients to allow

TABLE III

Reliability Coefficients

	I.L.A.M. Group A		R.A.M. Group A		W.K.C.S.	
	r Between Split-Halves	Corrected by Spearman-Brown Prophesy Formula	r Between Split-Halves	Corrected by Spearman-Brown Prophesy Formula	r Between Split-Halves	Corrected by Spearman-Brown Prophesy Formula
1. Musical Memory, Drake	.868	.93	.840	.91	.739	.85
2. Interval Disc., Drake	.586	.74	.720	.83	.276	.43
3. Retentivity, Drake	.624	.76	.578	.73	.363	.53
4. Cadence, Lowery	.375	.64			.134	.23
5. Tonal Sequence, Schoen	.013	.025				
6. Quality, Kwalwasser-Dykema	.500	.66			.248	.39
7. Melodic Taste, Kwalwasser-Dykema	.440	.61			.254	.40
8. Pitch, Kwalwasser-Dykema	.245	.39				
9. Tonal Memory, Kwalwasser-Dykema	.382	.55			.395	.57
10. Tonal Movement, Kwalwasser-Dykema	.738	.85			.575	.73
11. Pitch, Seashore	.641	.72			.727	.84
12. Intensity, Seashore	.739	.85			.795	.88
13. Tonal Memory, Seashore	.901	.94			.752	.86
14. Time, Seashore	.515	.68			.537	.70
15. Rhythm, Seashore	.499	.68			.314	.48
16. Rhythm, Kwalwasser-Dykema					.105	.19
17. Consonance, Seashore					.176	.30

for dividing the tests by halves, which are as high or higher than this, we have:—

Musical Memory
Pitch (Seashore)
Intensity (Seashore)
Tonal Memory (Seashore)

CONCLUSIONS

If a test is to be used for predictive purposes it must have both a high validity and a high reliability. When we consider the tests which have satisfactory coefficients in both validity and reliability we find only three, namely: Musical Memory, and Pitch and Tonal Memory by Seashore.

According to the analysis made in this research there is great room for improvement in tests of musical talent.

Only one out of twenty tests gave satisfactory results under all conditions to which they were subjected. The problem should not be considered solved because some tests have been constructed and a name given to them but this work should serve as the basis upon which to build more exact measuring instruments in the future.

REFERENCES

1. Billroth, T., *Wer Ist Musikalisch?* Berlin, 1896 (p. 164).
2. Binet, A., *Die Neuen Gedanken Uber Das Schulkind.* (Dtsch. Bearb. v. Anschutz, G. and Ruttmann, W. J.) Leipzig, 1912. (p. 199).
3. Copp, E. F., "Musical Ability," *J. of Heredity*, 1916, 7.
4. Crzellitzer, "Zur Methodik der Untersuchung auf Vererbung gustiger Eigenschaften, *Zsch. f. ang. Psych.*, 1911, 3.
5. Drake, R. M., "Four New Tests of Musical Talent," *J. Applied Psy.*, April, 1933, pp. 136-147.
6. Kohler, W., "Akustische Untersuchungen," *Zsch. f. Psych.*, 1910-1915, 54, 58, 64, 72.
7. Matthay, T., *The Act of Touch.* Longmans, Green, London, 1926 (p. 43).
8. Stumpf, C., *Tonpsychologie.* Band 2. Leipzig, 1890 (p. 157).
9. Wundt, W., "Eine Peplik C. Stumpf," *Ph. Studien*, 7 (p. 93).

A STUDY OF FACTORS CONTRIBUTING TO SCHOOL CITIZENSHIP[1]

M. E. BROOM
San Diego, California, State Teachers College
and
JOSEPHINE DOUGLAS BRAMKAMP
El Centro, California

The Problem

The approach to this study required first determining what characterized good school citizenship and poor school citizenship in the opinion of teachers and supervisors in a selected group of public schools. The problem of this investigation was to determine the contribution of various combinations of temperament and personality traits and of mental ability to school citizenship.

The Procedure

In order to secure a subject group for the study, the second writer solicited the co-operation of the teachers of all pupils in grades seven and eight in the rural schools of Imperial County, California. These persons also aided in the formulation of a citizenship traits rating scale which will be given hereafter. The subject group included one hundred twenty-two pupils, fifty-six boys and sixty-six girls, who responded to all of the tests of mental ability and of temperament and personality traits which were administered. Thirteen of these subjects were

[1] A report of a part of a master's thesis at the University of Southern California which was prepared by the second writer. The present manuscript was prepared by the first writer, who acted as thesis chairman for the original study.

Mexicans, nine were Japanese, and the remainder were Americans with various European nationalities represented in their heredity. Sixty-two of these pupils were in grade eight, and sixty, in grade seven.

The citizenship ratings were given by the classroom teachers. The tests were all administered by two supervisors who had had definite training in educational measurements and who had had previous experience in conducting school achievement surveys. The pupil responses to the tests were scored by three trained persons and this scoring was checked during the second scoring by the second writer to insure accuracy. The statistical computations were computed twice with the aid of statistical tables and calculating machines.

The Measuring Instruments

The citizenship traits rating chart is reproduced separately. Each positive and negative term was defined to prevent the teachers from making inaccurate ratings due to misinterpretation. These definitions were explained in detail by the second writer to each teacher prior to the giving of the ratings. In situations where more than one teacher had an opportunity to know of the child's school citizenship (the schools included from two to five classrooms), the additional teacher or teachers aided in preparing the pupil's citizenship rating.

A number of tests were administered. These will be considered here separately. Mental ability was tested by the Terman Group Test of Mental Ability, Form A. Symonds says that this test has proved its usefulness.[2] Odell says that it is more often rated as the best test for grades seven through twelve than is any other group test of mental ability.[3]

Three measures of temperament and personality traits

[2] P. M. Symonds, *Measurement in Secondary Education.* New York: The Macmillan Company, 1927. Page 29.

[3] C. W. Odell, *Educational Measurement in High School.* New York: The Century Company, 1930. Page 405.

Name, Surname first _____ Date _____

_____SCHOOL

TRAITS AND ATTITUDE RATING

Date of birth_____Grade_____Teacher_____

POSITIVE TRAITS	Above			AVERAGE	Below			NEGATIVE TRAITS
	Somewhat	Decidedly	Extremely		Somewhat	Decidedly	Extremely	
	7	6	5	4	3	2	1	
1. Civic minded								1. Self centered
2. Punctual								2. Dilatory
3. Enthusiastic								3. Indifferent
4. Mentally industrious								4. Mentally lazy
5. Self-confident								5a. Overconfident 5b. Lacking self-confidence
6. Courteous								6. Boorish
7. Co-operative								7. Self-seeking
8. Persevering								8. Unsteady
9. Reliable								9. Unreliable
10. Sportsmanlike								10. Unfair
11. Accurate								11. Careless
12. Mentally alert								12. Mentally slow
13. Logical								13. Scatterbrained
14. Adaptable								14. Obstinate
15. Purposeful								15. Aimless
16. Emotionally poised								16. Unstable
17. Creative								17. Lack instinct of workmanship
18. Sensitive to true values								18. Superficial
19. Independent								19. Dependent

Place a check in ink in the square which best expressed in your opinion the degree of the traits indicated.

were used. The first was the Pressey X-O Test for Investigating the Emotions, Form B. This test is composed of three sub-divisions: Test I, which aims to discover moral judgment; Test II, which aims to detect certain anxiety tendencies; and Test III, a measure of definite likes. In Test II one finds represented types of pathological attitudes, such as the paranoid or suspicious, the neurotic, the self-conscious, the melancholic, and the hypochondriacal. Freeman says that this test is the most ambitious attempt yet made to work out a test of the emotions.[4] Each sub-division of the test and the total test yield a total affectivity score and an idiosyncrasy score.

The MacNitt Psychological Interview yields a measure of extroversion-introversion. MacNitt says that the test is diagnostic.[5] Freyd has given an excellent discussion of the meaning and the characteristics of extroversion and introversion.[6] The material, however, is too extensive to reproduce here.

The Woodworth-Mathews Personal Data Sheet is a questionnaire comprising a list of seventy-five questions concerning habits of thought, emotional response, social adjustments, and volitional controls. It is more or less generally assumed that this instrument somehow gives an indication of emotional stability, but Fleming and Fleming say that in spite of all of the studies which have been made there are no data which clearly indicate what a score on the questionnaire or any adaptation of it may mean.[7]

[4] F. N. Freeman, *Mental Tests.* Boston: Houghton, Mifflin Company, 1926. Pages 208-217.

[5] R. D. MacNitt, *Manual for Interpreting the Psychological Interview.* Superior, Wisconsin: The author. Page 1.

[6] M. Freyd, "Introverts and Extroverts." *Psychological Review,* 31 (1924): 74-75, 78-79, 84-87.

[7] E. G. Fleming and C. W. Fleming, "The Validity of the Mathews' Revision of the Woodworth Personal Data Questionnaire." *Journal of Abnormal and Social Psychology,* 23 (1929): 500-505.

THE FINDINGS

One hundred twenty-two pupils' responses to the instruments listed in the paragraphs above are summarized here. The Terman mental age scores ranged between 10 years, 11 months and 18 years, 4 months, with a mean of 13 years, 0 months, and a standard deviation of 1 year, 4 months. The Pressey total affectivity score range varied between 56 and 334. The mean of this distribution was 165.1; the standard deviation, 48.5. The range of the Pressey idiosyncrasy scores varied between 27 and 62, and the mean was 48.5. The standard deviation describing the frequency variability was 6.9. The Personal Data Sheet responses ranged between 0 and 43. The mean was 13.4, and the standard deviation was 10.0. The range of the measures yielded by the MacNitt Psychological Interview included scores between 85 and 149. The mean was 117.5, and the standard deviation, 13.4. The range of the ratings for citizenship varied between 29 and 130, with a mean of 86.3, and a standard deviation of 21.9.

The quantitative evaluation of the contributions of mental ability and of the several measures of temperament and personality traits in securing high citizenship ratings are shown in Tables I to IV inclusive. Table I presents the data involving mental age and Pressey total affectivity (t) scores. Mental age above the average seems to be more significant in securing high citizenship ratings than is total emotional affectivity below the average. Apparently emotional stability and high mental ability are characteristics possessed by the better citizens.

Table II is concerned with the contributions of the Pressey idiosyncrasy scores (d) and of mental ability to citizenship. The indication is again that brighter pupils are better citizens, and among these brighter pupils those who have few emotional idiosyncrasies are

the better citizens. Among the groups, pupils below the average in mental ability who have the less normal emotionality are the better citizens. The findings in Table II with this exception agree with those for Table I.

TABLE I

A Quantitative Evaluation of the Contribution of Mental Age and of Emotional Affectivity to Citizenship. N is 122.

Personality Score		Mental age: Below the median (Md 159.7)	Mental age: Above the median (Md 159.7)
Above the median		N= 31 Range= 41-116 M= 79.7 S.D.= 20.0	N= 30 Range= 42-130 M= 91.5 S.D.= 20.8
Below the median		N= 30 Range= 29-115 M= 81.1 S.D.= 19.9	N= 31 Range= 37-130 M= 92.9 S.D.= 23.1

Table III presents the relationship between extroversion and introversion, teamed with mental ability and high citizenship or low citizenship. Among those pupils above the average in mental ability the introverts are the better citizens. In the group below the average in mental ability there is very little difference in the citizenship rating between the introverts and the extroverts. However, the difference indicates that the introverts are better citizens than the extroverts. As in Tables I and II, mental ability is more important in earning a high citizenship rating than is the trait of personality or temperament.

Table IV presents data in close agreement with that of Table I. In the groups both above and below the average in mental ability those pupils possessing the more normal personalities are better citizens than those indi-

TABLE II

A Quantitative Evaluation of the Contribution of Mental Age and of Emotional Stability to Citizenship. N is 122.

Personality Score	Mental age: Below the median	Mental age: Md 13-4	Mental age: Above the median
Above the median	N= 36 Range= 41-114 M= 81.4 S.D.= 20.5		N= 25 Range= 37-130 M= 88.1 S.D.= 22.0
Md 48.5			
Below the median	N= 25 Range= 29-115 M= 78.9 S.D.= 19.5		N= 36 Range= 42-125 M= 95.1 S.D.= 20.2

viduals who are less stable. The brighter pupils, too, are the better citizens in this grouping.

Summary and Conclusions

In determining the relationship of mental age and of personality and temperament traits to citizenship, the evidence seems to indicate that the brighter pupils are the better citizens. Among the pupils above the average in mental ability those with the more normal personalities are better citizens than the bright group of less stable emotional temperament. This is also true of the

TABLE III

A Quantitative Evaluation of the Contribution of Mental Age and of Extroversion-Introversion to Citizenship. N is 122.

		Mental age: Below the median	Mental age: Md 13-4	Mental age: Above the median
MacNitt Score	Above the median	N= 32 Range= 29-116 M= 80.4 S.D.= 19.8		N= 29 Range= 37-130 M= 93.1 S.D.= 23.6
	Md 116.8			
	Below the median	N= 29 Range= 42-115 M= 80.4 S.D.= 20.4		N= 32 Range= 42-130 M= 88.3 S.D.= 21.3

TABLE IV

A Quantitative Evaluation of the Contribution of Mental Age and of Emotional Stability to Citizenship. N is 122.

		Mental age: Below the median	Mental age: Md 13-4	Mental age: Above the median
Personal Data	Above the median	N= 37 Range= 41-116 M= 81.5 S.D.= 20.8		N= 24 Range= 37-130 M= 89.5 S.D.= 26.8
	Md 12.2			
	Below the median	N= 24 Range= 29-115 M= 82.8 S.D.= 17.1		N= 37 Range= 51-127 M= 94.0 S.D.= 18.1

groups below the average in mental ability, with one exception, the Pressey d score. In that test the group of pupils of poor mental ability with the less stable emotionality appears to include the better citizens. In the group above the average in mental ability and introvert, there is found higher citizenship than in the group above the average in mental ability and extrovert. In the group below the average in mental ability, citizenship is markedly similar throughout the entire group, regardless of extroversion and introversion. In the four tables as a whole the results indicate that emotional stability is a factor of doubtful significance, although the advantage in citizenship is with the more emotionally stable persons. To what degree emotional stability functions it is difficult to say. Insofar as citizenship is concerned the best combination of traits seems to be mental ability above the average, low emotional affectivity, emotional stability, and introversion.

It is difficult to draw positive conclusions from the findings of a study in which the subject group is so highly selected as is the case in the present instance. Before final judgment is passed on this problem similar research should be carried out in urban school systems and particularly in junior high schools. Such studies should involve larger numbers of cases, if possible, in order to insure the validity of the findings.

STUDY OF ABILITY TO WAKE AT A SPECIFIED TIME

KATHARINE T. OMWAKE and MARGARET LORANZ
Agnes Scott College

Many people claim to be human alarm clocks, able to wake at a specified time merely by reminding themselves of this time as they drop off to sleep. Such individuals say that when starting on a trip and leaving in the night or early morning they never find it necessary to depend on an alarm clock but wake spontaneously at the desired time. We hear wonderful tales of nurses who sleep soundly but are able to wake just in time to administer medicine no matter at what hour or at what intervals the medicine must be given. The ability to do this appears greatly exaggerated. Often it seems nothing more than the habit of waking at a particular time. Forgetting failures and remembering only the successful instances help magnify the stories told.

It has been suggested that the ability to sleep with one eye open consists in entering the sleeping state in which most sensory thresholds are raised, with the sensory threshold to one stimulus at somewhere near its normal waking level. Thus the nurse wakes at the slightest move of her patient but sleeps peacefully while the train thunders by. This waking in response to a definite if faint sensory stimulus is easier to explain than the situation of waking at a pre-determined time. In the latter case even though a time sense be postulated, the exact nature of the stimulus to waking is unknown.

A number of attempts have been made to experiment with people who claim to be able to wake at a set time.

Bond[1], although interested primarily in the mechanism of waking rather than in the accuracy of waking, reported waking within fifteen minutes of the desired time on each of the three occasions on which he gave definite figures. Hall[2] made one hundred night trials on himself, setting himself a waking time in the early morning. He reported exact successes in 18 per cent of the trials. In 53 per cent he was not more than fifteen minutes wrong; in 75 per cent not more than thirty minutes wrong; and in 81 per cent not more than forty-five minutes wrong. He concluded that a time sense exists, or that there is "no lack of evidence that a subconscious measurement of the lapse of time takes place."

Using a group instead of a single subject, the following experiment was undertaken to discover to what extent people actually can wake at an appointed hour during the night. The period of the study covered four weeks, the first and fourth weeks serving as controls, the second and third as experimental weeks. Twenty college girls served as subjects.

In a preliminary questionnaire ten of the subjects said that on the basis of their previous experience they thought they could waken themselves at any specified hour without the aid of an alarm clock. Ten had no confidence at all in their own ability to wake without an alarm.

As it was recognized that going to sleep with the intention of waking at an unusual time would disturb sleep, it was felt that a breaking in or control period would be desirable. In the first week an attempt was made to ascertain the hours of spontaneous waking. Each subject was instructed to place beside her pillow a flashlight, watch, pencil and pad of paper, and to record the

[1] Bond, N. B., *The Psychology of Waking*, Journal of Social and Abnormal Psychology, July-September, 1929, pp. 226-230.

[2] Hall, W. Winslow, *The Time Sense*, Journal of Mental Science, 1927, pp. 421-428.

time immediately if she woke. She was told not to try to wake but to keep an accurate record of the time of spontaneous waking and to make the record then—not in the morning. The necessity for accuracy and care was stressed; the subjects conscientiously tried to follow directions and there is no reason to believe that any but honest results were reported. Exactly the same procedure was followed the fourth week.

Every day of the second and third weeks each girl was notified at what hour to try to wake up that particular night. She was directed to concentrate on waking at the specified time, and to "will" as strongly as possible to wake at that time. She retired as usual with flashlight, watch, pad and pencil, prepared to record the exact minute of waking. If she woke several times she was to make a record of each time. The times specified were so scattered over the night as to have each hour represented several times during the two weeks, the times ranging from 12:30 to 6:15.

The chief weakness in the experimental procedure was the inability to motivate strongly and evenly, night after night. In spite of conscientious endeavor to follow directions the subjects reported difficulty in "willing" to wake. Several said they thought they would wake if they had a stronger motive and asked to be allowed to set an alarm clock, saying they were sure they would wake in time to prevent disturbing the whole dormitory. Although the weakness of motivation was recognized as a serious drawback this solution did not seem feasible.

After a month's experimentation we have little conclusive evidence. Of the twenty subjects only ten were at all successful and eight of these belonged to the group who felt sure they could wake. Three degrees of success were recognized; waking exactly at the time specified; within fifteen minutes of the time; and within half an hour of the time. The latter is admittedly not very close but appears worth considering particularly, as was usu-

ally the case, when it was the only waking during the night.

Table I shows the records of the ten subjects, with an average of 49 per cent partial successes, a marked tendency toward correct wakings.

TABLE I

Per Cent Successful Wakings in Fourteen Trials. Group A (Successful Group)

Subjects	Exactly on time	Within 15 min. of time	Within 30 min. of time
A	14%	21%	36%
B	7	21	29
C	0	29	43
D	7	21	50
E	0	14	29
F	21	50	86
G	7	29	43
H	7	57	78
I	36	36	36
J	0	21	57
	—	—	—
Average	10%	30%	49%

This tendency appears more evident when comparison is made with the record of Group B, composed of the ten less successful subjects. Of this group one woke on the dot once; there were five successes within fifteen minutes of the time and seven within thirty minutes. Five of these seven successes, including the only complete success, were contributed by the two subjects who stated in the preliminary questionnaire that they believed they could wake at will. Their success, while less than that of the eight others who had confidence in their ability to wake, was decidedly greater than that of eight of the ten not professing that ability. Two subjects, A and D, who before the experiment said that they did not think they could wake, were moderately successful.

TABLE 2

Per Cent Successful Wakings in Fourteen Trials. Group B (Unsuccessful Group)

Subjects	Exactly on time	Within 15 min. of time	Within 30 min. of time
K	0%	0%	0%
L	0	14	14
M	7	14	21
N	0	0	0
O	0	0	0
P	0	0	0
Q	0	0	0
R	0	0	0
S	0	0	7
T	0	7	7
	—	—	—
Average	1%	4%	5%

The earlier the hour the fewer the correct wakings. For example, at 5:30 all ten subjects in the successful group woke within half an hour of the stated time; while at 1:00 or 2:00 results were much less satisfactory. In the group of ten there were on the average 36 per cent successes when the specified hour was before 2:30; 42 per cent successes between 2:30 and 4:30; and 70 per cent successes after 4:30. As sleep became lighter towards morning it evidently became easier to wake. These results compare very favorably with those obtained by Bond and Hall who included in their studies only the early morning hours, obtaining a higher rate of success.

A knowledge of all wakings is necessary to interpret success. For example, subject I woke just five times, each time on the minute set. Subject F, on the other hand, with twelve successes to her credit woke about twice a night. Her sleep was evidently always disturbed, however, for her record for the control weeks shows almost as many wakings as during the experiment.

Comparison of the number of wakings of the successful and unsuccessful groups reveals a significant difference. This difference appears only in the weeks of the experiment (second and third). During the control weeks (first and fourth) when no effort was made to wake the two groups were about the same in their sleeping habits. During the weeks of experimental waking, however, a difference appears which may explain the success of one group and the failure of the other. Both groups

TABLE 3

Number of Wakings per Week.

Subjects	First week	Second week	Third week	Fourth week
Successful group	43	72	68	27
Unsuccessful group	49	36	26	26
Total	92	108	94	53

found sleep disturbed at the beginning of the first week by the very fact of going to bed with flashlight, watch, pad and pencil, with attention centered on recording wakings. With the unsuccessful group the disturbance to sleep gradually became less and each succeeding week showed fewer wakings. This group actually woke less often during the experiment proper than during the first control week. Not only did they not wake at the specified time, they did not wake at all.

The successful group had sleep more disturbed during the experiment, waking almost twice as often during the experimental weeks as during the controls. This fact is in line with the theory that going to sleep with the intention of waking at some unusual time means less profound sleep with the threshold to waking stimuli lowered. Also, perhaps, the greater number of wakings led to some chance successes.

In spite of the admitted weakness in motivation during the experiment we feel that our results are at least suggestive. The experiment seems to justify the following conclusions.

1. Ability to wake at a specified time is a special ability, exercised more effectively by some than others.
2. Half of the subjects woke within thirty minutes of the stated time in 49 per cent of the trials.
3. One subject, if she woke at all, did so at the exact time.
4. The group of subjects who thought they could wake did definitely better than the group who did not think they could.
5. Attempting to wake at a set time disturbs sleep.
6. The per cent of wakings on the hours after 4:30 was much larger than on the hours before 4:30.

THE COMPARATIVE CONSTANCY OF STANFORD-BINET AND HERRING-BINET IQ'S

CLAUDE L. NEMZEK
University of Minnesota

Recently Carroll and Hollingworth[1] reported a study in which 52 gifted children had two Stanford-Binet IQ ratings and two Herring-Binet IQ ratings, the intervals between testings being one year. The purpose of the present discussion is not only to show the comparative constancy of the IQ's obtained from these two particular tests but also to point out a method which may be used, for instance, in comparing the constancy of IQ's (a) derived from individual and group mental tests, (b) obtained by different examiners, (c) of boys as contrasted with girls, (d) according to levels of intelligence, (e) of various chronological age groups, and (f) of so-called normal individuals versus other groups.

The r between Stanford-Binet test-retest IQ's was .68±.051; for the Herring-Binet, .73±.044. Ranges of IQ's, means, probable errors of the means, standard deviations, probable errors of the standard deviations, and probable errors of measurement were calculated as shown in Table I. The difference in Stanford-Binet means was 5.6±1.024, D/P.E.$_D$ being 5.47; the difference in Herring-Binet means was 3.0±.935, D/P.E.$_D$ being 3.21.

[1] Herbert A. Carroll and Leta S. Hollingworth, "The Systematic Error of Herring-Binet in Rating Gifted Children," *Journal of Educational Psychology*, 21: 1-11, January, 1930, No. 1.

TABLE I

Tests	Ranges of IQ's	Mean	P.E.$_{(Av.)}$	S.D.	P.E.$_{\sigma}$	P.E.$_{(M)}$
S.-B.$_1$	133 to 190	152.3	1.3	13.85	.92	5.21
S.-B.$_2$	128 to 188	157.9	1.26	13.45	.89	
H.-B.$_1$	108 to 174	135.8	1.3	13.8	.91	4.75
H.-B.$_2$	116 to 173	138.8	1.24	13.3	.88	

TABLE II

Measures of Changes	Gains	Losses	All Changes Algebraically	All Changes Regardless of Signs
Number of Cases	33	18	52*	52
Range of Change	2 to 32	1 to 21	—21 to 32	0 to 32
Median Change	8.93	6.00	6.33±1.323	8.33±.607
Q_1 of Change	6.75	3.63	—3.75	5.00
Q_3 of Change	13.75	8.75	11.50	12.00
Q of Change	3.50	2.56	7.63±.833	3.50±.382
Mean Change	10.82	6.22	4.71±.945	9.02±.613
S.D. of Change	6.86	4.41	10.10+.668	6.55±.433

* One case did not gain or lose.

TABLE III

Measures of Changes	Gains	Losses	All Changes Algebraically	All Changes Regardless of Signs
Number of Cases	33	18	52*	52
Range of Change	1 to 22	2 to 19	—19 to 22	0 to 22
Median Change	10.63	6.00	5.00±1.359	8.80±.856
Q_1 of Change	5.63	3.75	—4.00	4.33
Q_3 of Change	14.44	13.50	11.67	14.20
Q of Change	4.41	4.88	7.84±.856	4.94±.539
Mean Change	10.06	7.72	3.71±.938	9.06±.532
S.D. of Change	5.54	5.40	10.03±.663	5.69±.376

* One case did not gain or lose.

TABLE IV

Measures Compared	Stanford-Binet	Herring-Binet	D	P.E.$_D$	D/P.E.$_D$
r (Test-Retest IQ's)	.68± .051	.73± .044	.05	.067*	.746
Median Alg. Change	6.33±1.323	5.00±1.359	1.33	1.763**	.754
P_{85} to P_{15} Alg.***	15.25±1.672	15.67±1.718	.42	2.228	.189
Q of Alg. Change	7.63± .833	7.84± .856	.21	1.110	.189
Mean Alg. Change	4.71± .945	3.71± .938	1.00	1.238	.808
S.D. of Alg. Change	10.10± .668	10.03± .663	.07	.932	.075
Median Change Reg.	8.33± .607	8.80± .856	.47	1.114	.422
P_{85} to P_{15} Reg.***	7.00± .768	9.87±1.082	2.87	1.408	2.038
Q of Change Reg.	3.50± .382	4.94± .539	1.44	.701	2.054
Mean Change Reg.	9.02± .613	9.06± .532	.04	.864	.046
S.D. of Change Reg.	6.55± .433	5.69± .376	.86	.568	1.514

* The uncorrelated data formula was used in calculating this probable error of a difference.

** The correlated data formula was used in calculating all of the other probable errors of differences in this table. The r between the changes from the first Stanford-Binet test to the second and the changes from the first Herring-Binet test to the second was .136±.092, when the changes were considered algebraically; —.134±.092, when the changes were considered regardless of signs.

*** Harry V. Masters and C. C. Upshall, "Table of Probable Errors for Certain Inter-percentile Ranges," *Journal of Educational Psychology*, 23:287-290, April, 1932, No. 4.

In Table II are shown various measures of changes concerning the Stanford-Binet data. Similar calculations for the Herring-Binet are presented in Table III. Using these data the writer made a statistical comparison, concluding from the facts contained in Table IV that there are no significant differences in the constancy of the Stanford-Binet and Herring-Binet IQ's of the 52 gifted children concerned.

INTELLIGENCE AS A CAUSAL FACTOR DETERMINING DIFFERENCES IN SUSTAINED ATTENTION IN PRESCHOOL CHILDREN

HELEN S. SHACTER
Northwestern University

Differences in the attentive activity of children have been noted by many observers and commentators, and various factors have been suggested as being causal elements therefor.

The chronological age of the subjects has been suggested as explanatory of differences in sustained attention. This view, holding that the older the child the longer the period of sustained attention, has been expressed by Andrus (1), Bertrand (3), Bott (6), Davidson (13), Herring and Koch (18), Inskeep (21), Van Alstyne (32), and Woodrow (34).

Again, it has been indicated that the sex of the subjects determines their attention-span, but although sex has been frequently suggested as an influential factor, there has been disagreement as to which sex exhibits the greater attention-span. According to Bertrand (3), Bridges (8, 9) and Van Alstyne (32), girls sustain attention longer than do boys; Goodenough (12) indicated that boys tend to perseverate more than do girls, and Herring and Koch (18) observed less sustained interest in girls than in boys.

Fundamental differences in personality development have also been offered as the essentials explaining differences in attention behavior by Brown (10), studying preschool children, and by Hovey (19) who used high school pupils as subjects.

At least one writer, Goodenough (16), has concluded that only children do not attend for as long periods of time as do children having siblings.

Perhaps the most frequent explanation offered to account for the sustaining of attention has been intelligence, a positive relationship being held to exist between the degree of intelligence manifested and the length of the attention-span. Binet (5), Ordahl (25), Sollier (27), Terman (29), and Tredgold (31) have all expressed a conviction that the feebleminded may be distinguished from the normal by their inability to sustain attention. That attentive behavior and intelligence are closely related has been emphasized by a number of writers who considered children of various ages as their subjects. Burt (11) in 1909 offered the conclusion that attention is indispensable in the manifestation of intelligence, after securing teachers' opinions concerning the mental ability of the subjects of a test which he designed as a measure of attention. McQueen (23) in 1917 stated that the higher the intelligence of an individual the more capable he is of intense concentration, using the judgments of three teachers as acceptable data on which to formulate this theory. Lobsein (22) in 1919 also reported a high correlation between intelligence and attention, using as his basis certain unstandardized intelligence tests of his own.

There have been published results of some investigations where the measure of intelligence was objective rather than subjective and therefore was a more dependable means of comparison. In Hughes' study (20) made in 1926 high school students were the subjects and the Terman Group Test of Mental Ability was used, its results being correlated with the Hughes Rating Scale for Individual Capacities, Attitudes, and Interests. A coefficient of .34 was reported to have been found between intelligence and control of attention. Cushing (12) investigated a "perseverative tendency" in young

children in 1929 and found a correlation of .58 between the Merrill-Palmer results and the perseverative score. Brown (10) in 1930 used the Stanford-Binet Scale, resulting intelligence quotients being compared with initial, final and best mean scores for a series of reaction-time experiments performed as measures of attention; coefficients of correlation resulting from these comparisons were, respectively, .30, .60, and .60. The subjects used by Brown were preschool children, as were those studied by Nelson (24) in 1931. Nelson obtained a .10 relationship between a persistence score and Binet results, and a .01 correspondence between persistence and Merrill-Palmer measures.

The varied opinions of the many writers upon the subject of attention have only too often been theories without supporting experimental evidence. The observational studies reported of child activities have been made under such diverse conditions that comparative results may scarcely be considered valid. The desirability of fact rather than of conjecture as a basis for educational and social procedure need scarcely be stressed.

The writer has suggested elsewhere that a possible reason for the lack of scientific data concerning the attention process in children has been the difficulty of measuring such activity, and has proposed a method for measuring the sustained attention of young children (26).

The use of this method with a group of thirty-six preschool children resulted in evidence indicating that children of three, four, and five years of age show no significant differences in the time of sustained attention. The subjects of the investigation were all enrolled in either the Prekindergarten or the Kindergarten of the Elementary School of Chicago Teachers College, and the groups at the three-year levels noted were equal in number. Each subject engaged in six different situations, at which the time of sustained attention was recorded.

Three of these situations were designated as "simple," and involved merely repetitive movements, offering no possibilities of variation; the remaining three situations were designated as "complex," involving the same repetitive movements but permitting a diversified performance which presented possibilities of varied approaches and methods of attack. The experimental materials utilized conformed with the findings of Wagoner (33), of Bridges (8, 9), and of Herring and Koch (18) in investigations concerning the preferences of young children for play materials, as indicated by spontaneous choices from a wide selection of play equipment. None of the activities set up as an experimental situation had an obvious termination but, rather, lent itself to repetitive manipulation which tended to perpetuate itself.

Examination of the individual records obtained indicated a consistent brief or long-sustained period of attention and suggested that some causal element or elements remain to be designated as accounting for the varied attention-spans noted. It was obvious that the average attention-spans of the subjects in the two types of situations presented—simple and complex—differed: the average periods of sustained attention ranged from eight to twelve minutes, *depending upon the complexity of the activity engaged in.*

This suggested a relationship existent between attention-span and intelligence; the present study is an attempt to evaluate intelligence as a causal factor of differences in attentive behavior.

To ascertain the degree of possible relationship existent, it was necessary to obtain an objective measure of the intelligence of the subjects of the experiment. No one scale or test available seemed wholly satisfactory from all aspects; where one places great emphasis upon the acquisition of language, and thus penalizes the child hesitant in the use of words, another capitalizes on the manipulative inclinations of young children and dis-

criminates against the child who is awkward and not adept with his hands. Terman (29) pointed out in writing of the Binet-Simon Scale that

> . . . No single test used alone will determine accurately the general level of intelligence. A great many tests are required, and for two reasons: (1) because intelligence has so many aspects; and (2) in order to overcome the accidental influences of training or environment.[1]

For the investigation here reported the composite result of three widely varying types of tests was considered as the basis of comparison of intelligence with attention-span. The procedure utilized a verbal scale, a performance scale, and a picture-pointing test, *viz.*, the Stanford Revision of the Binet-Simon Scale (29), the Merrill-Palmer Performance Scale (28), and the Detroit Kindergarten Test (2). There were thus obtained three intelligence quotients for each subject, which were averaged to provide a composite IQ for each individual examined; this composite is what will hereafter be referred to as "intelligence quotient" in the presentation of data. By this procedure the final intelligence measure considered gave no undue advantage to the verbalist type over the shy or slow-of-speech child, nor did it favor the manually dextrous subject and discriminate against the child less clever or experienced with his hands.

Analysis of the data on intelligence reveals that the distribution for each of the three age-levels examined shows considerable homogeneity, the four-year-old group averaging only a slightly higher intelligence quotient than the three-year-olds, 110.9 and 108, respectively, and the latter averaging only a slightly higher intelligence quotient than the five-year-olds, whose average was 105.

[1] L. M. Terman, *The Measurement of Intelligence*, p. 77.

The group of subjects as a whole is seen to be somewhat selected insofar as mental ability is concerned. The range of intelligence quotients is from 84 to 135, with an average of 108, and an average deviation of 7.4. A little better than forty per cent of the subjects scored at or higher than 110 in the composite results of the intelligence measurements.

With a division of the group according to sex there is to be noted a fairly even distribution for two-thirds of each section: twelve boys and twelve girls score between 100 and 115 in the intelligence test results. Of the six remaining girls, three fell below this range and three scored above it; of the six boys, two fell below and four scored above it. The lowest individual measurement of the entire group was that made by a five-year-old girl; the highest individual intelligence quotient was scored by a four-year-old boy.

It has been indicated that a very slight difference was found between the three groups of subjects of the investigation—three, four, and five-year-old children—in the average time of sustained attention, as measured by the experimental procedure utilized. Within each group, however, there appeared marked quantitative differences in the attention process of the individual children. Considering both the simple and complex situations of the experiment, when the three approaches of each type of activity were averaged, a wide range was found to exist in the attention-span at each age-level. In the simple situations the records of the three-year-old children ranged from 2′ 19″ to 21′ 51″; that of the four-year-olds varied between 2′ 22″ and 18′ 24″; the five-year-old group had attention-spans from 2′ 25″ to 20′ 18″. Variations were also evident in the complex situations, the range here being 4′ 26″ to 28′ 25″ for the three-year-olds, 6′ 7″ to 32′ 34″ for the four-year-olds, and 4′ 33″ to 17′ 16″ for the five-year-olds.

The average attention-spans in the simple situations were 8′ 22″, 8′ 15″, and 9′ 22″ for the three, four, and five-year-old subjects, respectively; in the complex situations, the average attention-spans were 11′ 2″, 12′ 55″, and 11′ 14″, respectively.

With these average attention-spans were correlated the average composite intelligence quotients obtained. The number of cases was such that large probable errors were inevitable, yet definite tendencies could be reliably indicated. The rank-difference method was used for computation and the resulting coefficients were transmuted into product-moment coefficients by means of Table XX,[2] in Garrett (14). These indications of relationship appear in Table I where the same data for the group as a whole are presented.

TABLE I

Correlations Between Composite Intelligence Quotients and Average Attention-Spans in Simple and in Complex Situations, for Each Age-Level, and for the Entire Group.

	Simple Situations		Complex Situations	
Age	*r*	*PE*	*r*	*PE*
3	.49	.15	.50	.15
4	.31	.18	.56	.14
5	—.04	2.03	.24	.19
3, 4, 5	.12	.12	.33	.10

The lack of marked relationship as shown by these correlations is striking, only the highest coefficient, .56, showing a somewhat significant correspondence between the two factors under consideration. It is not surprising to find a consistently higher degree of relationship between the intelligence measurements and the time of sustained attention for the complex situations as com-

[2] H. E. Garrett, *Statistics in Psychology and Education*, p. 192.

pared with the intelligence quotients and the attention-spans for the simple situations.

It is to be noted that the correlation drops off with the simple situations as age increases; in a smaller degree the same thing occurs with the complex situations. This may indicate a very definite relationship between the significance of the material to the subjects and the length of time they will attend to it: simple material intrigues the very young child, but as age increases simple material loses its charm and only the more complex material is capable of offering a challenge. It is interesting to speculate whether the six-year-old would sustain attention with the same materials as were utilized in the experimental approach here described.

When the three age-groups of the investigation are considered as one distribution of preschool subjects the same difference in correspondence between complexity of material and length of attention-span is seen. Yet even for the complex situations where the relationship is more marked the coefficient is so low, .33 and its probable error comparatively so high, .10, that it may not be interpreted as evidence of a causal relationship. The low positive relationship found to exist between the sustaining of attention and mental ability indicates that differences in intelligence cannot be used to account for differences in attentive activity in individual children.

It seems to the writer significant that when *estimates* of intelligence were used by investigators as a basis of comparison of mental ability and attention, the relationship reported stated intelligence to be the essential factor in determining attention, whereas, in those studies in which an *objective measure* of intelligence was used as a basis of comparison a decidedly lower degree of correspondence was invariably revealed. While these several results are not strictly comparable with those of the present study—indeed, Brown (10) is the only investigator noted whose work purported to study attention ob-

jectively—they are still sufficiently relevant to make the data of interest in connection with the investigation here reported.

The hypothesis that there is a definite causal relationship between intelligence and sustained attention is not supported by the results of the present investigation. It would be presumptuous to make any attempt at a generalization from the evidence of the number of cases examined in this study, but certainly it may justifiably be suggested from the results that some factor other than mental ability is involved and must be sought if an explanation is to be attempted of the so obviously varying manifestations of attention in individuals.

BIBLIOGRAPHY

1. Andrus R. An inventory of the habits of children from two to five years. New York: Bureau of Publications, Teachers College, Columbia University, 1928. Pp. 51.

2. Baker, H. J. and Kaufman, H. J. Detroit kindergarten test. Yonkers: World Book Co., 1925.

3. Bertrand, F. L. Contribution a'l'étude psychogénétique de l'attention. *L'Année psychol.*, 1925, 26, 155-158.

4. Binet, A. Attention et adaptation. *L'Année psychol.*, 1899, 6, 248-404.

5. Binet, A. Les indées modernes sur les enfants. Paris: E. Flammarion, 1909. Pp. 346.

6. Bott, H. Observation of play activities in a nursery school. *Genet. Psych. Monog.*, 1928, 4, 44-88.

7. Bridges, K. M. B. Critical notes on mental tests for children of preschool age. *Ped. Sem.*, 1927, 34, 38-44.

8. Bridges, K. M. B. Occupational interests of three-year-old children. *Ped Sem.*, 1927, 34, 415-423.

9. Bridges, K. M. B. Occupational interests of four-year-old children. *J. Genet. Psychol.*, 1929, 36, 551-569.

10. Brown, M. Continuous reaction as a measure of attention. *Child Develop.*, 1930, 1, 255-291.

11. Burt, C. Experimental tests of general intelligence. *Brit. J. Psychol.*, 1909, 3, 153-157.

12. Cushing, A. M. A perseverative tendency in preschool children. *Archiv. Psychol.*, 1929, No. 108.

13. Davidson, H. P. An experimental study of bright, average, and dull children at the four-year mental level. *Genet. Psychol. Monog.*, 1931, 9, 125-289.

14. Garrett, H. E. Statistics in psychology and education. New York: Longmans, Green & Co., 1930. Pp. 317.

15. Gesell, A. The mental growth of the preschool child. New York: MacMillan Co., 1925. Pp. 447.

16. Goodenough, F. L. The emotional behavior of young children during mental tests. *Jour. Juv. Research*, 1929, 13, 204-219.

17. Hallowell, D. K. Mental tests for preschool children. *Psychol. Clinic*, 1927, 16, 235-276.

18. Herring, A. and Koch, H. L. A study of some factors influencing the interest span of preschool children. *J. Genet. Psychol.*, 1930, 38, 249-275.

19. Hovey, H. B. Effects of general distraction on the higher thought processes. *Amer. J. Psychol.*, 1928, 40, 585-591.

20. Hughes, W. H. Relation of intelligence to trait characteristics. *J. Educ. Psychol.*, 1926, 17, 482-494.

21. Inskeep, A. D. Child adjustment in relation to growth and development. New York: D. Appleton & Co., 1930. Pp. 427.

22. Lobsein, M. Prüfung der Aufmerksamkeit im Kindern mit der Münsterbergsher Schlittenmethode. *Zeit. f. angewandte Psychol.*, 1919, 14, 40-74.

23. McQueen, E. N. The distribution of attention. *Brit. J. Psychol.*, Monog. Suppl. 5, 1917. Pp. 142.

24. Nelson, J. F. Personality and intelligence. New York: Teachers College Bureau of Publications, Columbia Univ., 1931. Pp. 62.

25. Ordahl, E. O. and G. Qualitative differences between levels of intelligence in feebleminded children. *J. Psycho-Asthenics*, Monog. Suppl., 1, 1915. Pp. 49.

26. Shacter, H. S. A method for measuring the sustained attention of preschool children. *Jour. of Genet. Psychol.*, 1933, 42, 339-371.

27. Sollier, P. A. Psychologie de l'idiot et de l'imbecile. Paris: F. Alcon, 1901. Pp. 236.

28. Stutsman, R. Mental measurement of preschool children. Yonkers: World Book Co., 1931. Pp. 365.

29. Terman, L. M. The measurement of intelligence. Boston: Houghton Mifflin Co., 1916. Pp. 362.

30. Terman, L. M. The physical and mental traits of gifted children. In Twenty-third Yearbook, National Society for Study of Education. Public School Publishing Co., 1924. Pp. 155-167.

31. Tredgold, A. F. Mental deficiency. New York: Wm. Wood & Co., 1920. Pp. 525.

32. Van Alstyne, D. Play behavior and choice of play materials of preschool children. Chicago: Univ. of Chicago Press, 1932. Pp. 100.

33. Wagoner, L. The constructive ability of young children. *Univ. Iowa Stud.:* Stud. in Child Welfare, 1925, 3, 3-55.

34. Woodrow, H. Brightness and dullness in children. Philadelphia: J. B. Lippincott Co., 1919. Pp. 322.

35. Wooley, H. T. The validity of standards of mental measurement in young childhood. *School and Society*, 1925, 21, 476-482.

BOOK REVIEWS

WOODWORTH, ROBERT L. *Adjustment and Mastery.* Baltimore: The Williams & Wilkins Company in cooperation with The Century of Progress Exposition, 1933. 137 pp.

In this small volume Professor Woodworth considers the major divisions of psychology in their relation to adjustment to and mastery of life situations. Sensation and perception are largely adjustment; learning is largely mastery. Mental set in reading becomes "context-adjustment," and emotion is organic adjustment. Individual differences, intelligence tests, the interaction of heredity and environment, come in for brief mention. Maladjustments and crime are discussed with the emphasis upon prevention through getting at the fundamental causes. Adjustment and mastery are largely to be achieved through teamwork which is the subject and title of the concluding chapter. The author says, "Good teamwork is the highest form of human activity. The great artist or scientist, though he be a lone worker, actually keeps in touch more or less directly with the work of his fellows. His work ties in with theirs. Group activity, integrating the acts of individuals towards a common goal, is not only the most productive type of human endeavor, but the most satisfying to the individual. To make the world a better place for the individual, then, we have to improve the organization of group activity. We have to take account of the motives and desires that are common to men, and also the differing abilities of men, and so organize our social activities, in work and play, as to provide for each individual an outlet for his energies."

This goal, the author believes, is a distant one, but one that must be reached if we are to receive more than a dubious benefit from the progress of the physical sciences; if we are to make "progress" good for us.

DUNLAP, KNIGHT, AND GILL, ROBERT SUTHERLAND. *The Dramatic Personality of Jesus.* Baltimore, The Williams & Wilkins Company, 1933. Pp. vi + 186.

Whatever may be the case as to the historicity of Jesus, which the authors allow for reasons of argumentative convenience, it is

largely as a dramatic personality that Jesus has been of importance to the world. It is pointed out that a dramatic personality should be consistent. It "must be that of a person who could logically have lived in the place and at the time dramatically assigned to it." Obviously the Jesus of the gospels is not entirely such a personality. There are inconsistencies which must be ironed out if he is to emerge a sound personality. From a consideration of his racial and social background, his youth and adult life as it must have been lived in its particular time and place, there emerges a personality that is sound and consistent. Following chapters deal with the ethical teachings, ecclesiastical views, views on family and state relationships, and the theology of Jesus. The product of his time, he is too intelligent to be limited by it, and his teachings are the logical resultant of such a personality in such a society.

In the chapter on *Jesus as a Charlatan* the miracles are reduced to those of healing, which in view of present day faith healing of functional disorders become no miracles at all. The other miracles and the impossible cases of healing, e. g., of lepers, are ascribed to exaggeration. But that Jesus believed himself to have miraculous powers is not out of keeping with his personality. Here we see him perhaps most clearly as a product of his time. The real nature of disease was not understood, and healing by contact and verbal formula was commonly accepted. The writers consider this matter at length, since a dramatic personality must be *real*, and the miracles if accepted at face value, make Jesus a fictitious or impossible character.

This dramatic personality is not the Jesus of the gospels nor is it the Jesus of the churches, but it is for the authors the Jesus who might have been, and the Jesus who, unlike the other Jesuses, did not really fail.

AMOS C. ANDERSON,
Ohio University.

GILBERT, L. C., *An Experimental Investigation of Eye Movements in Learning to Spell Words.* Psychological Monographs, No. 196. Psychological Review Co., Princeton. 1932. viii + 81 p. $1.50.

The eye movements made by good and poor spellers in studying spelling were compared with each other at every grade level from grade III through the senior high school. The better spellers demonstrated more efficient oculomotor and perceptual efficiency at each level. In general the study of spelling was found to involve one or more cross-passages of fixations along the word. The for-

ward progression of fixations was usually interrupted by regressions. With increase in maturity came an improvement in efficiency of study habits, a change from minute examination of the word to apprehension of letter combinations in larger units. The good spellers revealed more efficient oculomotor performance at each level of maturity. The author might have stressed the fact that eye-movement habits are highly symptomatic of orderliness and degree of efficiency in the central processes of perception.

On the basis of changes in efficiency from grade to grade, rather definite conclusions have been drawn concerning development of proficiency in eye-movement habits. These deductions cannot be accepted as adequate or justified since sampling errors (only three to four subjects in each group) and selective factors (elimination of inefficient scholars with increase in grade level) were not satisfactorily controlled.

The second part of the study contains an excellent analysis of methods of attack in studying spelling and the implications of these findings for teaching procedures. On the whole the monograph is an important preliminary study in the field. Both the psychologist and the class room teacher will profit from its contents.

MILES A. TINKER,
University of Minnesota.

NEW BOOKS AND PAMPHLETS RECEIVED

Books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

An Evaluation of School Health Procedures. RAYMOND FRANZEN. Monograph V. American Child Health Association, 450 Seventh Avenue, New York City, 1933. $1.15. 127 pp.

Child Psychology. ARTHUR T. JERSILD. Prentice-Hall, New York City, 1933. $3.00. 462 pp.

Case Studies in the Psychopathology of Crime. BEN KARPMAN. Mimeoform Press, Washington, D. C. 1042 pp.

College and Life. M. E. BENNETT. McGraw-Hill Book Company, New York City, 1933. $2.75. 456 pp.

Comparison of Delinquent and Non-Delinquent Boys by Objective Measures of Personality. MARJORIE E. BARCOCK. University of Hawaii, Honolulu, Hawaii. 73 pp.

The First Two Years: Personality Manifestations, Vol. III. MARY M. SHIRLEY. University of Minnesota Press, University of Minnesota, Minneapolis, Minn. $2.50. 288 pp.

Human Studies in Hawaii. S. D. PORTEUS. Proceedings and Lectures, School of Oriental and Pacific Affairs, University of Hawaii, 1932. 32 pp.

Introductory Acoustics. GEORGE WALTER STEWART. D. Van Nostrand Company, New York. 200 pp.

Medicina de Trabajo e Higiene Industrial, Vol. V, No. 17-18, 1933. Instituto de Reeducacion Profesional, Madrid, Spain. 183 pp.

Psychology and the New Education. S. L. PRESSEY. Harper & Brothers, Publishers, New York, 1933. 594 pp.

University of Iowa Studies, Iowa City, Iowa:

Bladder Control in Infancy and Early Childhood. HJALMAR FLETCHER SCOE. Vol. V, No. 4, June, 1933. 83 pp.

The Companionships of Preschool Children. ELIZABETH PLEGER HAGMAN. Vol. VII, No. 4, 1933. 69 pp.

Learning in Preschool and Orphanage Children. KATHERINE ELIOTT ROBERTS. Vol. VII, No. 3, 1933. 94 pp.

A Study of Some Factors in Form Board Accomplishments of Preschool Children. Harold Manville Skeels. Vol. VII, No. 2, 1933. 147 pp.

Directory of American Psychological Periodicals

(Continued)

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Pedagogical Seminary and Journal of Genetic Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages, approximately. Edited by Walter V. Bingham. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Philosophy—New York: 515 West 116th Street. Subscription $4.00. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush, and H. W. Schneider.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Edited by Robert Saudek (London) and Ernest Seeman (Durham, N. C.). Founded 1932.

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monograph—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300-400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $5. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Indianapolis: The C. E. Pauley & Co., Inc. Subscription $5.00. 400 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1917.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

(Continued on Third Cover)